

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> C12N 15/12, C07K 14/705, A61K 38/17, G01N 33/68 // C12N 15/11, A61K 31/70	<b>A1</b>	<b>(11) International Publication Number:</b> WO 95/22610 <b>(43) International Publication Date:</b> 24 August 1995 (24.08.95)
<b>(21) International Application Number:</b> PCT/US95/02044 <b>(22) International Filing Date:</b> 15 February 1995 (15.02.95) <b>(30) Priority Data:</b> 08/199,776 18 February 1994 (18.02.94) US <b>(71) Applicant (for all designated States except US):</b> BRIGHAM AND WOMEN'S HOSPITAL, INC. [US/US]; 75 Francis Street, Boston, MA 02115 (US). <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> BRENNER, Michael, B. [US/US]; 11 Gardner Road, Brookline, MA 02146 (US). PARKER, Christina, M. [US/US]; 21 Beacon Heights Drive, Newton Centre, MA 02159 (US). <b>(74) Agent:</b> PLUMER, Elizabeth, R.; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210 (US).		<b>(81) Designated States:</b> AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MW, MX, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SI, SK, TJ, TT, UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ, UG).  <b>Published</b> <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
<b>(54) Title:</b> NOVEL INTEGRIN ALPHA SUBUNIT  <b>(57) Abstract</b>  The present invention relates to a novel integrin $\alpha^E$ subunit and its functional equivalents. The invention further includes pharmaceutical compositions containing the isolated peptides, oligonucleotides encoding the peptides, vectors containing the oligonucleotides, and cell lines transfected with the vectors.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

Novel Integrin Alpha Subunit  
Government Support

This invention was made in part with government support under grant numbers R01CA47724, R29GM49342 and K11AI00903 from the National Institutes of Health.

Field of the Invention

This invention relates to a novel integrin alpha chain; functionally-equivalent peptide fragments and analogs thereof; oligonucleotides encoding the peptide fragments and analogs; vectors containing and cell lines expressing the novel peptides; and methods for using the peptide fragments, analogs and oligonucleotides.

Background of the Invention

The integrin mediated adhesive interactions of cells with other cells and between cells and the extracellular matrix are believed to play critical roles in a wide variety of processes including, for example, modulation of the immune system, regulation of developmental processes and tumor progression and metastasis. These molecules also transduce information from the extracellular to the intracellular environment through poorly understood signalling mechanisms. The integrins represent one of the best characterized superfamilies of adhesion receptors. Integrins are glycoprotein heterodimers which contain a non-covalently associated  $\alpha$  and  $\beta$  subunit. Integrin subunits are transmembrane proteins which contain an extracellular domain for interacting with an extracellular matrix or cellular component, a transmembrane domain spanning the cell membrane and a cytoplasmic domain for interacting with one or more cytoskeletal components.

There are fourteen known  $\alpha$  subunits and eight known  $\beta$  subunits which can pair to form at least twenty different integrin molecules. Several distinct integrin  $\alpha$  chains are capable of pairing with one type of  $\beta$  chain to form a  $\beta$  chain subfamily. Thus, for example, the  $\beta_1$  subfamily includes seven members (also known as the VLA proteins:  $\alpha^1\beta_1$ - $\alpha^7\beta_1$ ); the  $\beta_2$  subfamily includes three members (the leukocyte cell adhesion molecules or LeuCAMs:  $\alpha^L\beta_2$  or LFA-1,  $\alpha^M\beta_2$  or Mac-1 and  $\alpha^X\beta_2$  or p150,95) and the  $\beta_3$  subfamily includes two members ( $\alpha^V\beta_3$ ,  $\alpha^{IX}\beta_3$ ). In some instances, an  $\alpha$  chain may pair with more than one  $\beta$  chain, e.g.,  $\alpha^4$  can pair with  $\beta_1$  or  $\beta_7$ .

The integrin  $\alpha$  chains have in common a seven-fold repeated amino acid motif, of which the last three or four motifs include divalent cation binding sites. All known  $\alpha$  chains have been divided into one of two structural groups on the basis of amino acid sequence homology and the presence or absence of two structural features (described below).

The first group of  $\alpha$  chains contains a proteolytic cleavage site located in the extracellular domain, proximal to the transmembrane region. Post-translational cleavage of the  $\alpha$  chain precursor yields two fragments which (with one exception) remain associated by a disulfide linkage. The smaller fragment includes a short portion of the extracellular domain, the transmembrane and the cytoplasmic domains. The larger fragment contains the major portion of the  $\alpha$  chain extracellular domain. This group of post-translationally cleaved integrin  $\alpha$  subunits includes  $\alpha^3$ ,  $\alpha^4$ ,  $\alpha^5$ ,  $\alpha^6$  (formerly called  $\alpha^2$ ),  $\alpha^v$  and  $\alpha^{Ib}$ , although the  $\alpha^4$  molecule is a more distant member of the group since it is less similar to the other cleaved integrin  $\alpha$  subunits based upon homology analysis, is cleaved near its mid-point to yield two fragments of nearly equal size and further, because the fragments are not disulfide linked (Teixido, J. et al., (1992) J. Biol. Chem. 267, 1786-1791; Rubio, M. et al., (1992) Eur. J. Immunol. 22, 1099-1102).

Members of the second group of integrin  $\alpha$  subunits do not include the above-described proteolytic cleavage site. Moreover, the second group of  $\alpha$  subunits is characterized by the presence of an additional region known as the "I" (inserted) domain. Homologous I domains have been identified in complement factors B and C2, von Willebrand's factor, cartilage matrix glycoprotein and collagen type VI.

The importance of integrins with respect to modulation of the immune system is illustrated by the condition, leukocyte adhesion deficiency (LAD), a disorder that is characterized by profound immunodeficiency. Individuals afflicted with LAD are unable to express the  $\beta_2$  integrin subfamily (Hogg, N. (1989) Immunol. Today 10, 111-114). Thus, while it has been known for some time that integrins and other adhesion molecules function in immune system modulation, e.g., by playing a role in the adhesion of peripheral lymphocytes to endothelium and in homing to lymph nodes. However, relatively little is known about the molecules that function in the

mucosal immune system, a subset of the general immune system which includes the lymphocytes which populate the gastrointestinal, genito-urinary and respiratory tracts, and the mammary glands. In particular, little is known about the molecules which function in mucosal lymphocyte homing. (see, Cepek, K. et al., (1993) J. Immunol. 150, 3459-3470 and references cited therein).

Recently, we described a novel integrin heterodimer that is expressed on intra-epithelial T lymphocytes (iIEL), i.e., the population of T lymphocytes located along the baso-lateral surfaces of the epithelial cells which line the mucosa, adjacent to the epithelial cell basement membrane. (Parker, C.M. et al., (1992) Proc. Natl. Acad. Sci. U.S.A. 89, 1924-1928). Originally defined by an antibody which recognizes the human mucosal lymphocyte 1 antigen (HML-1), the novel integrin is present on >90% of intestinal IEL (iIEL) and on approximately 40% of lamina propria T lymphocytes (which lie between the epithelial basement membrane and the muscularis mucosae) (Cerf-Bensussan, N. et al., (1987) Eur. J. Immunol. 17, 1279-1285). The HML-1 antigen contains a novel  $\alpha$  chain (designated  $\alpha^E$ , for "epithelial associated") associated with a  $\beta$  chain (Parker, C.M. et al., (1992) Proc. Natl. Acad. Sci. U.S.A. 89, 1924-1928). Although the HML-1  $\beta$  chain has been cloned (Yuan, Q.A. et al., (1990) Int. Immunol. 2, 1097-1108; Erle, D.J. et al., (1991) J. Biol. Chem. 266, 11009-11016), little is known about the primary structure of the  $\alpha^E$  chain.

Cloning of the  $\alpha^E$  chain has proven to be problematic because of the difficulty in obtaining an adequate number of intra-epithelial cells from which the  $\alpha^E$  chain could be purified and sequenced. In addition, the relatively large length of the  $\alpha^E$  gene has hampered cloning efforts because of the propensity to lose the 5' portion of relatively long genes during cDNA synthesis.

#### Summary of the Invention

The cDNA sequence and derived amino acid sequence (Sequence I.D. Nos. 1 and 2, Genbank Accession Number L25851) for the HML-1  $\alpha$  chain are disclosed herein. The  $\alpha^E$  chain was purified from a hairy cell leukemia spleen lysate and the N-termini of the purified protein fragments (25kDa and 150 kDa) were sequenced to prepare degenerate oligonucleotide probes for screening a cDNA library. Surprisingly, in view of its resemblance to other integrin  $\alpha$  subunits in overall amino acid sequence,  $\alpha^E$  contains a region of 55

amino acids (referred to herein as the "X" (extra) domain or Sequence I.D. No. 4, encoded by Sequence I.D. No. 3) located N-terminal to the I domain. This region is not present in any other integrin.

The X domain contains two unique structural features: (1) an internal proteolytic cleavage site (between amino acids 159 and 160 of Sequence I.D. No. 1) followed by (2) a highly charged region of 18 consecutive amino acids (amino acids 162-179). Upon in vivo proteolytic cleavage of the  $\alpha^E$  chain, two fragments are formed: a 26 kD fragment having a C-terminus which contains the 34 amino acids of the X domain N-terminus and a 150 kD fragment having an N-terminus which contains the 21 amino acids of the X domain C-terminus. The 21 amino acid portion of the X domain (referred to hereinafter as Sequence I.D. No. 5) includes the highly charged region of 18 consecutive amino acids. The 34 amino acid portion of the X domain N-terminus is referred to hereinafter as Sequence I.D. No. 6. These unique structural features of the X domain, coupled with the discovery that high levels of the mRNA for  $\alpha^E$  (and for  $\beta_7$ , with which it associates to form the integrin  $\alpha^E\beta_7$ ) are restricted to mucosal lymphocytes, suggested to us that the  $\alpha^E$  subunit likely plays a unique role in the localization and/or site specific functions of intra-epithelial T-lymphocytes.

One aspect of the invention is directed to isolated peptides which inhibit the in vivo and in vitro function of the  $\alpha^E$  subunit. The peptides have sequences which are related to, or derived from, the amino acid sequence of the above-described X domain (Sequence I.D. No. 4), e.g., Sequence I.D. Nos. 5, 6 and functionally equivalent peptide analogs of the foregoing peptides.

According to another aspect of the invention, a method for selecting a functionally equivalent peptide analog of Sequence I.D. No. 4 is provided. The method includes providing a peptide analog of Sequence I.D. No. 4 and determining whether the peptide analog inhibits adhesion between a human mucosal lymphocyte-1 antigen and an epithelial cell in vitro. Preferably, the peptide analogs are between about four and about twenty amino acids in length. More preferably, the peptide analogs are between about four and about ten amino acids in length. Exemplary peptide analogs are disclosed in Sequence I.D. Nos. 8 through 25, inclusive.

According to yet another aspect of the invention, a method for

screening a molecular library to identify lead compounds which inhibit the in vivo activity of the integrin  $\alpha^E$  chain is provided. The method includes determining whether the molecular library contains a compound which inhibits adhesion between a human mucosal lymphocyte-1 antigen and an epithelial cell in vitro. Also provided is a competitive binding assay method for identifying lead compounds which mimic the ligand binding site of the integrin  $\alpha^E$  chain. The method involves determining whether the library contains a molecule which competitively inhibits the binding of the  $\alpha^E$  chain (or a functionally equivalent peptide fragment or analog thereof) to an antibody which specifically recognizes the ligand binding site of the  $\alpha^E$  chain.

According to another aspect of the invention, a pharmaceutical composition is provided. The composition includes a therapeutically effective amount of one or more of the above-identified isolated peptides (e.g., Sequence I.D. No. 4, a fragment or functionally equivalent peptide analog thereof) and a pharmaceutically acceptable carrier therefor. Optionally, the composition further includes the  $\beta_7$  chain or a portion thereof in association with the above-identified isolated peptide. Alternatively, the pharmaceutical composition includes a therapeutically effective amount of one or more isolated oligonucleotides and a pharmaceutically acceptable carrier therefor.

According to still another aspect of the invention, a support having a biologically active surface which exhibits cell attachment activity is provided. The support includes a surface to which is attached one or more of the above-disclosed peptides. Exemplary supports include a prosthesis device (e.g., a vascular graft, a percutaneous device) and an affinity matrix (e.g., for isolating the ligand(s) of the  $\alpha^E$  subunit).

According to another aspect of the invention, a method for isolating a cell surface ligand of the integrin  $\alpha^E$  chain is provided. The method involves coupling Sequence I.D. No. 4, or a functionally equivalent fragment or peptide analog thereof (e.g., Sequence I.D. Nos. 5-6 and 8-25), to an inert support and isolating the cell surface ligand using affinity chromatography.

According to yet another aspect of the invention, antibodies to the isolated peptides or isolated oligonucleotides are provided.

The antibodies are useful for blocking a functional activity of intra-epithelial T lymphocytes, such as an in vivo functional activity (e.g., localization of the T lymphocytes) or an in vitro functional activity (e.g., adhesion of intra-epithelial T lymphocytes to an epithelial cell monolayer as determined in an adhesion assay). Accordingly, the antibodies are useful as reagents in screening assays to identify lead compounds that are present in molecularly diverse libraries or other mixtures.

According to yet another aspect of the invention, an isolated oligonucleotide is provided. The isolated oligonucleotide encodes a peptide selected from the group consisting of Sequence I.D. Nos. 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 and 25. In a particularly preferred embodiment, the isolated oligonucleotide comprises Sequence I.D. No. 7, an oligonucleotide which encodes Sequence I.D. No. 6.

Also provided is an isolated oligonucleotide that is capable of hybridizing under stringent conditions (defined below) to the nucleotide sequence residing between about position 555 and about position 656 of Sequence I.D. No. 1. This region of the  $\alpha^E$  cDNA encodes Sequence I.D. No. 6 (i.e., the relatively uncharged proteolytic cleavage fragment of the X domain). In a preferred embodiment, the isolated oligonucleotide is at least about 80% homologous with the above-recited nucleotide region of Sequence I.D. No. 1. More preferably, the isolated oligonucleotide is 100% homologous with the nucleotide sequence residing between position 555 and position 656 (referred to hereinafter as Sequence I.D. No. 7) inclusive of the  $\alpha^E$  cDNA.

According to still another aspect of the invention, an antisense oligonucleotide capable of hybridizing under stringent conditions to the above-described isolated oligonucleotide is provided. The antisense oligonucleotide is capable of hybridizing to a unique fragment (defined below) of the naturally-occurring DNA or mRNA encoding the  $\alpha^E$  subunit. Accordingly, delivery of an antisense oligonucleotide to intra-epithelial lymphocytes in vivo inhibits localization of the lymphocytes by base-pairing with the DNA (or RNA) encoding a unique fragment of the  $\alpha^E$  nucleic acid, thereby preventing transcription (or translation) of the  $\alpha^E$  subunit.

According to yet other aspects of the invention, a recombinant



expression vector comprising at least one strand of the above-disclosed isolated oligonucleotide and a cell line transfected with the recombinant expression vector are provided. Preferably, the oligonucleotide is operatively joined to at least one regulatory sequence, for example, a promoter or enhancer sequence. Suitable cell lines include mammalian cells; bacterial cells; insect cells and various yeast strains.

These and other aspects of the invention, as well as various advantages and utilities, will be more apparent with reference to the detailed description of the preferred embodiments and in the accompanying drawings.

#### Brief Description of the Drawings

Figure 1 schematically illustrates the composite sequence for the integrin  $\alpha^E$  subunit cDNA clone and the sequencing strategy used to obtain the clone;

Figure 2 illustrates the similarity of the  $\alpha^E$  subunit with other integrin  $\alpha$  subunits based upon generation of an integrin  $\alpha$  subunit homology tree;

Figure 3 illustrates the similarity of the  $\alpha^E$  subunit with other integrin  $\alpha$  subunits based upon the determination of the percent similarity of the  $\alpha^E$  amino acid sequence to those of other integrin  $\alpha$  subunits;

Figure 4 illustrates the immunoprecipitation of the  $\alpha^E$  subunit isolated from the HML-1 antigen with antiserum generated to the C-terminal peptide deduced from the  $\alpha^E$  cDNA; and

Figure 5A shows that the  $\beta$ , probe detected a 3 kb mRNA that hybridized abundantly with RNA derived from cultured iIEL and hairy cell spleen.

Figure 5B shows that multiple tissues and cell lines which showed the presence of the 4.5 kb  $\alpha^E$  transcript also showed the 3 kb  $\beta$ , transcript, including lung, spleen, thymus, prostate, ovary, small intestine, colon, and peripheral blood leukocytes.

Figure 5C shows that multiple tissues and cell lines which showed the presence of the 4.5 kb  $\alpha^E$  transcript also showed the 3 kb  $\beta$ , transcript, including fresh peripheral blood monocytes and T cells, tonsillar B cells.

Figure 5D shows that multiple tissues and cell lines which showed the presence of the 4.5 kb  $\alpha^E$  transcript also showed the 3 kb  $\beta$ , transcript, including PEER, and also in MOLT 13 and MOLT 17

at barely detectable levels.

Figure 5E shows that  $\alpha^E$  transcripts were detected in the absence of detectable levels of  $\beta$ , in RNA samples derived from Va2, RD and HeLa and 16E6.A5.

Figure 6 shows a schematic representation of the two  $\alpha^E$  genomic clones and the knockout construct.

Figure 7 shows the strategy to identify  $\alpha^E$  homologous recombination.

### Detailed Description of the Invention

#### I. Introduction to the Preferred Embodiments

The nucleic acid sequence and deduced amino acid sequence (Sequence I.D. Nos. 1 and 2) for a cDNA encoding the human integrin  $\alpha^E$  subunit are disclosed herein. The cDNA clone was obtained by: (1) isolating the  $\alpha^E$  subunit of the HML-1 antigen from a hairy cell leukemia spleen lysate and sequencing the N-terminus of the purified protein fragments (25 kDa and 150 kDa); (2) preparing degenerate oligonucleotide probes based upon the N-terminal amino acid sequences and screening a lambda Zap II cDNA library to obtain individual clones encoding portions of the  $\alpha^E$  subunit; (3) sequencing the individual cDNA clones and preparing a composite cDNA sequence based upon overlapping sequences of individual clones; and (4) deducing the amino acid sequence for the composite cDNA. Each of these steps is described in detail in the Examples.

Two features distinguish the  $\alpha^E$  subunit from previously reported alpha subunits. First, upon proteolytic cleavage, the  $\alpha^E$  subunit is cleaved into a smaller (25 kDa) fragment containing Sequence I.D. No. 6 and a larger (150 kDa) fragment containing Sequence I.D. No. 5. The sequence for the smaller fragment shows sequence homology to the known integrin alpha chain amino termini, suggesting that unlike most other integrin alpha subunits which are cleaved proximal to the C-terminus or near the mid-point of the precursor polypeptide, the  $\alpha^E$  precursor is cleaved proximal to the N-terminus. Second, the N-terminal sequence for the larger (150 kDa) fragment contains a region of negatively charged residues (figure 1) which do not share sequence homology with any known integrin  $\alpha$  subunit sequence. As will be described in more detail below, these unique properties of the  $\alpha^E$  subunit result from the presence of an additional domain (the X domain) which contains an internal proteolytic cleavage site proximal to its N-terminus.

Degenerate oligonucleotide probes having sequences based upon the N-terminal amino acid sequences of the 150 and 25 kDa fragments, were used to screen a lambda Zap II cDNA library. The composite  $\alpha^E$  cDNA sequence, represented schematically in figure 1, was established by sequencing individual clones (depicted by thick black lines beneath the composite cDNA) and aligning the overlapping sequences. A comparison of the empirically-determined amino acid sequence for the HML-1  $\alpha^E$  subunit (figure 1, top line) and the amino acid sequence deduced from the cDNA sequence (figure 1, bottom line) demonstrated that the composite cDNA encodes the HML-1 antigen  $\alpha^E$  subunit. This result was confirmed by immunoprecipitation experiments and other methods of analysis (e.g., measuring biochemical features predicted from the  $\alpha^E$  cDNA, including molecular mass, proteolytic cleavage site, and the presence of N-linked glycosylation) (refer to the Examples).

The nucleotide sequence of the composite cDNA and the deduced amino acid sequence are disclosed in Sequence I.D. Nos. 1 and 2, respectively. Alignment of the deduced amino acid sequence with the published sequences of other I domain-containing human integrin alpha subunits and the generation of a homology tree demonstrated substantial sequence homology between the I domain-containing subunits (Figures 2 and 3). However, the  $\alpha^E$  subunit is a relatively distant member of the integrin I domain-containing alpha subunit family, based upon both overall sequence homology and the inclusion of the X domain (indicated in bold italics in Figure 1).

## II. The Unique X Domain

The nucleotide sequence and deduced amino acid sequence for the  $\alpha^E$  subunit cDNA are shown in Sequence I.D. Nos. 1 (including the signal sequence as indicated) and 2, respectively. As used herein, the X domain refers to the unique region of 55 amino acids located proximal to the N-terminus of the I domain in the  $\alpha^E$  subunit, i.e., amino acid position 126 through position 180 (encoded by nucleotide positions 555 through 719) inclusive of Sequence I.D. No. 1). Sequence I.D. No. 4 represents the deduced amino acid sequence of the X domain. Sequence I.D. No. 3 represents the oligonucleotide sequence encoding Sequence I.D. No. 4.

The X domain contains a proteolytic cleavage site, located between amino acids 159 and 160 of Sequence I.D. No.1, which distinguishes the  $\alpha^E$  subunit from all known integrin I domain-

containing alpha subunits. Cleavage of the X domain (Sequence I.D. No. 4) yields two fragments designated Sequence I.D. Nos. 5 and 6. Sequence I.D. No. 5 contains a stretch of eighteen consecutive charged residues which are located immediately C-terminal to the cleavage site. Accordingly, the nucleic acid sequence which encodes Sequence I.D. No. 5 is identical to the coding sequence which encodes the N-terminal portion of the 150 kDa  $\alpha^E$  chain.

In view of the unique primary structure of the X domain and our observation that high levels of  $\alpha^E$  mRNA are found primarily in mucosal lymphocytes, we believe that the X domain is integral to the adhesion of T lymphocytes to epithelial cells in vivo. Accordingly, the invention involves producing agents which modulate that adhesion and that are useful for treating autoimmune diseases that are characterized by lymphocyte accumulation at epithelial sites (e.g., ulcerative colitis, Crohn's disease, celiac disease, sarcoidosis, psoriasis, and the late phase component of asthma). Other agents also are useful for enhancing  $\alpha^E$  mediated adhesion, thereby permitting the design of more appropriate therapies for treating infectious diseases of epithelial sites (e.g., pulmonary tuberculosis, leprosy, cutaneous leishmaniasis, and parasitic or viral infections diseases of the intestinal tract) by increasing the function of mucosal intraepithelial lymphocytes or their localization to the epithelium.

According to one aspect of the invention, an isolated peptide that is capable of inhibiting the adhesion of  $\alpha^E\beta$ , to an epithelial cell in vitro also is provided. The isolated peptide has an amino acid sequence which is related to, or derived from, the amino acid sequence of the X domain (Sequence I.D. No. 4) of the  $\alpha^E$  subunit. The isolated peptide is selected from the group consisting of Sequence I.D. Nos. 4, 5, 6 and functionally equivalent peptide analogs of the foregoing peptides. In a particularly preferred embodiment, the isolated peptide is Sequence I.D. No. 4, 5 or 6.

As used herein in reference to a peptide, the term "isolated" refers to an expression product of an isolated oligonucleotide; a peptide which is isolated following cleavage from a larger polypeptide; or a peptide that is synthesized, e.g., using solution and/or solid phase peptide synthesis methods as disclosed in, for example, U.S. 5,120,830, the entire contents of which are incorporated herein by reference.

polypeptide; or a peptide that is synthesized, e.g., using solution and/or solid phase peptide synthesis methods as disclosed in, for example, U.S. 5,120,830, the entire contents of which are incorporated herein by reference.

As used herein, the term "peptide analog" refers to a peptide which shares a common structural feature with the molecule to which it is deemed to be an analog. A functionally equivalent peptide analog is a peptide analog which further shares a common functional activity (e.g., inhibiting the adhesion between an intra-epithelial T lymphocyte and an epithelial cell in vitro) with the molecule to which it is deemed an analog. Peptide analogs include unique fragments which are related to, or derived from, Sequence I.D. No. 4, polymers of Sequence I.D. No. 4, and polymers of unique fragments of Sequence I.D. No. 4. A "unique fragment" of a protein or nucleic acid sequence is a fragment which is not currently known to occur elsewhere in nature (except in allelic or allelomorphic variants). Unique fragments act as a "signature" of the gene or protein from which they are derived. A unique fragment will generally exceed 15 nucleotides or 5 amino acids. One of ordinary skill in the art can readily identify unique fragments by searching available computer databases of nucleic acid and protein sequences such as Genbank, (Los Alamos National Laboratories, USA), EMBL, or SWISS-PROT. A unique fragment is particularly useful, for example, in generating monoclonal antibodies or in screening genomic DNA or cDNA libraries.

It will be appreciated by those skilled in the art that various modifications of the foregoing peptide analogs can be made without departing from the essential nature of the invention. Accordingly, it is intended that peptides which include conservative substitutions (see, e.g., Table 2) and fusion proteins in which a peptide of the invention is coupled to a solid support (such as a polymeric bead), a carrier molecule (such as keyhole limpet hemocyanin), a toxin (such as ricin) or a reporter group (such as radiolabel or other tag), also are embraced within the teachings of the invention. Preferably, the peptide analogs contain between about four and about twenty amino acids. More preferably, the peptide analogs contain between about four and about ten amino acids. Exemplary peptide analogs which are fragments of Sequence I.D. Nos. 4 and 5 are provided in Table 1.

TABLE 1Peptide Analogs (fragments of Sequence I.D. Nos. 4 and 5)

EKEEEEDKEE	(Sequence I.D. No. 8)
KEEEDKKEE	(Sequence I.D. No. 9)
EEEDKKEEE	(Sequence I.D. No. 10)
EEEDKKEEED	(Sequence I.D. No. 11)
EEDKKEEED	(Sequence I.D. No. 12)
EDKKEEED	(Sequence I.D. No. 13)
DKEEED	(Sequence I.D. No. 14)
KEEED	(Sequence I.D. No. 15)
EEED	(Sequence I.D. No. 16)
ALEKEEEDK	(Sequence I.D. No. 17)
EEDKKEEED	(Sequence I.D. No. 18)
ALEKEEEDKKEE	(Sequence I.D. No. 19)
EKEEEDKKEEED	(Sequence I.D. No. 20)
ALEKEEEDKKEEED	(Sequence I.D. No. 21)

As used herein, "conservative amino acid substitution" refers to an amino acid substitution which does not alter the relative charge or size characteristics of the peptide in which the amino acid substitution is made. Conservative substitutions of amino acids include substitutions made amongst amino acids within the following groups: (a) MILV; (b) FYW; (c) KRH; (d) AG; (e) ST; (f) QN; and (g) ED. Exemplary peptide analogs of Sequence I.D. No. 5 which include conservative amino acid substitutions (e.g., E for D and/or K for N) are provided in Table 2.

TABLE 2Peptide Analogs having Conservative Amino Acid Substitutions

EKEEED	(Sequence I.D. No. 22)
DKDDDDDDDD	(Sequence I.D. No. 23)
EDEEEEEDE	(Sequence I.D. No. 24)
ENEEEDNEEED	(Sequence I.D. No. 25)

As used herein, the term "functionally equivalent peptide analog" refers to a peptide analog that is capable of inhibiting the binding of an intra-epithelial T lymphocyte to an epithelial cell in vitro. An in vitro adhesion assay (see, e.g., the adhesion assay provided in the Examples) is used as a screening assay to measure the ability of a molecule, e.g., a peptide analog, to

inhibit  $\alpha^E\beta_7$ -mediated adhesion between an intra-epithelial T lymphocyte and an epithelial monolayer in culture and is predictive of the ability of a molecule to inhibit the functional activity of the novel  $\alpha^E\beta_7$  subunit in vivo. For example, a "functionally equivalent peptide analog" of Sequence I.D. No. 4 includes fragments of Sequence I.D. No. 4, as well as peptide analogs of Sequence I.D. Nos. 4-6 which contain conservative amino acid substitutions, provided that the peptide fragments and analogs are capable of inhibiting adhesion of a human intra-epithelial lymphocyte to an epithelial cell.

According to yet another aspect of the invention, a method for selecting a functionally equivalent peptide analog of Sequence I.D. No. 4 is provided. The method involves providing a peptide analog of Sequence I.D. No. 4 and determining whether the peptide analog inhibits adhesion between a human mucosal lymphocyte-1 antigen and an epithelial cell in vitro. In a particularly preferred embodiment, the functionally equivalent peptide analogs are fragments of Sequence I.D. No. 4 containing between about four and about twelve amino acids and having about the same isoelectric point as Sequence I.D. No. 5 (see, e.g., Table 1, above).

According to yet another aspect of the invention, a method is provided for screening a molecular library to identify pharmaceutical lead molecules, i.e., molecules which inhibit the adhesion between an intra-epithelial T lymphocyte and an epithelial cell in vitro. A molecular library refers to a collection of structurally-diverse molecules and includes both recombinantly-produced and chemically-synthesized libraries. As used herein, "library molecule" refers to a molecule that is present in the molecular library. Accordingly, screening refers to the process by which library molecules are tested for the ability to inhibit the adhesion between an intra-epithelial T lymphocyte and an epithelial cell or the ligand binding activity of the  $\alpha^E\beta_7$  subunit. For example, the ability of a molecule to inhibit the binding of a mucosal T-lymphocyte (or the  $\alpha^E\beta_7$  subunit) to an epithelial cell in vitro can be used as a screening assay to identify lead compounds. Thus, a particularly preferred screening method involves determining whether the library molecule (or peptide analog) inhibits adhesion between a human mucosal lymphocyte-1 antigen and an epithelial cell in culture. Such adhesion assays

are well known in the art and are illustrated by the assay provided in the Examples.

Alternatively, antibodies which specifically recognize the ligand binding site of the  $\alpha^{\text{P}}\beta$ , integrin (i.e., that portion of the subunit which binds to an epithelial cell) can be used to screen combinatorial libraries (e.g., by contacting the library with the antibody and determining whether the library contains a molecule which competitively inhibits binding of the antibody to a peptide known to contain the ligand binding site). Such competitive binding assays also can be used to assess the relative affinity of the library molecule for its cognate (e.g., a ligand present on an epithelial cell). Antibodies to the ligand binding site also can be used to purify the integrin heterodimer or its subunits which bind to ligand expressed on the epithelial cell.

According to yet another aspect of the invention, methods and compositions for modulating the functional activity of the  $\alpha^{\text{P}}\beta$ , integrin in vivo are provided. One method involves administering a pharmaceutical composition containing a therapeutically effective amount of the isolated peptides and/or oligonucleotides of the invention. In general, the therapeutically effective amount is between about 1 ug and about 100 mg/kg. The preferred amount can be determined by one of ordinary skill in the art in accordance with standard practice for determining optimum dosage levels of the agent. The peptides (and/or oligonucleotides) are formulated into a pharmaceutical composition by combination with an appropriate pharmaceutically acceptable carrier. For example, the peptides may be used in the form of their pharmaceutically acceptable salts, or may be used alone or in appropriate association, as well as in combination with other pharmaceutically active compounds. The peptides may be formulated into preparations in solid, semisolid, liquid or gaseous form such as tablets, capsules, powders, granules, ointments, solutions, suppositories, and injections, in usual ways for oral, parenteral, or surgical administration. Exemplary pharmaceutically acceptable carriers are described in U.S. 5,211,657, the entire contents of which patent are incorporated herein by reference.

The invention also includes locally administering the composition as an implant or any part of it either alone or bound to a synthetic material, such as a vascular prosthesis.



Accordingly, a support (e.g., nitrocellulose, polyester, polyvinyl, polystyrene or a ceramic) having a biologically active surface which exhibits cell attachment activity also is provided. As used herein, a "biologically active surface which exhibits cell attachment activity" refers to a support to which is attached at least one of the above-identified isolated peptides. Attachment of the peptides confers upon the support surface the ability to attach cells, and in particular, the ability to attach epithelial cells. Methods for preparing a biologically active surface by coupling a peptide to an inert surface are disclosed in U.S. 5,211,657, the contents of which have been incorporated herein by reference. The peptides are coupled to plastic surfaces for in vitro use (e.g., for cell culture in which it is desirable to attach cells to a culture vesicle), to a prosthesis such as a vascular prosthesis or other artificial organ to make the synthetic materials more biocompatible and to allow the cells to adhere and grow or to an affinity matrix. Alternatively, the peptides are incorporated into a polymer during the polymerization process, rather than attached to the surface of a previously formed inert support.

According to yet another aspect of the invention, a method for isolating a ligand for the integrin  $\alpha^5\beta_1$  subunit from detergent extracts of cells or cell membranes is provided. The method includes (1) specifically adsorbing the ligand to an affinity matrix having Sequence I.D. No. 4 or a functionally equivalent peptide analog thereof coupled to the matrix to form a ligand-adsorbed affinity matrix, and (2) adding to the ligand-adsorbed affinity matrix a plurality of peptides selected from the group consisting of Sequence I.D. No. 4 and a functionally equivalent peptide analog of Sequence I.D. No. 5 to specifically elute the ligand from the affinity matrix. In a particularly preferred embodiment, the peptide coupled to the matrix is Sequence I.D. No. 4, 5 or 6.

As used herein, "affinity matrix" refers to an inert support to which molecules can be covalently attached in a manner such that the ligand binding portion of the molecule is exposed. The methods for coupling a molecule (e.g., a peptide) to an affinity matrix are determined by the nature of the functional groups present on the molecule and on the matrix (e.g., amine groups, carboxyl groups).

The peptide can be derivatized (according to standard procedures known in the art) to include additional functional groups to facilitate the coupling of the peptide to the matrix in a desired orientation. Typically, matrices are sold in an activated form for coupling to a specified class of functional groups. In such instances, instructions for coupling a molecule to the matrix are provided by the matrix manufacturer. Numerous types of coupling methods to a variety of inert supports are well known to those of ordinary skill in the art.

According to another aspect of the invention, an antibody specific for the  $\alpha^E$  subunit is provided. The antibodies are raised against the above-identified isolated peptides and/or fusion peptides using well known immunization techniques (see, e.g., the Examples: generation of an antibody to the synthetic C-terminal peptide predicted from the  $\alpha^E$  cDNA). Preferably, the antibody specifically recognizes an antigen selected from the group consisting of the X domain (Sequence I.D. No. 4) and fragments or equivalents of the X domain. In a particularly preferred embodiment, the peptides are selected from the group consisting of Sequence I.D. Nos. 4-6 and 8-21. The antibodies are useful in screening assays for identifying pharmaceutical lead compounds which inhibit the adhesion of an intra-epithelial T lymphocyte to an epithelial cell.

The antibodies also are useful in vivo for blocking cell adhesion and localization of intra-epithelial T lymphocytes and for example, targeting a toxin (e.g., ricin) or detectable agent (e.g., a radiolabel, a fluorescent label, an enzyme label) to cells (e.g., intra-epithelial lymphocytes) containing the  $\alpha^E$  subunit. Methods for coupling such toxins and/or agents to proteins and/or antibodies for in vivo and in vitro applications are disclosed in, for example, Killen and Lindstrom (1984), "Specific killing of lymphocytes that cause experimental Autoimmune Myasthenia Gravis by toxin-acetylcholine receptor conjugates", J. Immun. 133:1335; Jansen, F.K., et al. (1982), "Immunotoxins: Hybrid molecules combining high specificity and potent cytotoxicity", Immunolog. Rev. 62:185-216, the entire contents of which references are incorporated herein by reference. See also U.S. Patent Nos. 3,652,761; 4,478,946 and 4,554,088, the entire contents of which patents are incorporated herein by reference.

The invention also provides isolated oligonucleotides (e.g., Sequence I.D. No. 3) that encode the X domain (Sequence I.D. No. 4), its proteolytic fragments (Sequence I.D. Nos. 5 and 6) and the above-described functionally equivalent peptide analogs of the foregoing amino acid sequences (Sequence I.D. Nos. 8-25). As used herein, the term "isolated" in reference to an oligonucleotide, means an RNA or DNA polymer, portion of genomic nucleic acid, cDNA or synthetic nucleic acid which, by virtue of its origin or manipulation: (a) is not associated with all of a nucleic acid with which it is associated in nature (e.g., is present in a host cell as a portion of an expression vector); or (b) is linked to a nucleic acid or other chemical moiety other than that to which it is linked in nature; or (c) does not occur in nature. By "isolated" it is further meant a nucleic acid sequence: (i) amplified in vitro by, for example, the polymerase chain reaction (PCR); (ii) synthesized by, for example, chemical synthesis; (iii) recombinantly produced by cloning; or (iv) purified from a more complex molecule or from a mixture of molecules, such as by cleavage and size fractionation. Due to the degeneracy of the genetic code, many different oligonucleotide sequences can be identified which encode Sequence I.D. Nos. 4-6 and 8-25. Accordingly, the invention embraces oligonucleotides which encode the X domain (as well as its fragments) but which have nucleotide sequences which differ from the sequences of the naturally-occurring  $\alpha^E$  gene or from the cDNA disclosed in Sequence I.D. No. 1.

In addition to the foregoing oligonucleotides, the invention also provides an isolated oligonucleotide that is capable of hybridizing under stringent conditions to the nucleotide sequence residing between position 555 and position 719 inclusive of Sequence I.D. No. 1 (i.e., the region of the  $\alpha^E$  cDNA which encodes the X domain (Sequence I.D. No. 4)). As used herein, the phrase "hybridizing under stringent conditions" is a term of art which refers to the conditions of temperature and buffer concentration which will permit hybridization of a particular oligonucleotide or nucleic acid to its complementary sequence and not to non-complementary sequences. The exact conditions which constitute "stringent" conditions, depend upon the length of the nucleic acid sequence and the frequency of occurrence of subsets of that

sequence within other non-identical sequences. By varying hybridization conditions from a level of stringency at which no hybridization occurs to a level at which hybridization is first observed, one of ordinary skill in the art can, without undue experimentation, determine conditions which will allow a given sequence to hybridize only with identical sequences. Suitable ranges of such stringency conditions are described in Krause, M.H. and S.A. Aaronson, *Methods in Enzymology*, 200:546-556 (1991). Stringent hybridization conditions, depending upon the length and commonality of a sequence, may include hybridization conditions of from 30 to 60 degrees C and from 5x to 0.1x SSC. Highly stringent hybridization conditions may include hybridization at 45 degree C and 0.1 SSC. Less than stringent conditions are employed to isolate nucleic acid sequences which are substantially similar, allelic or homologous to any given sequence. In a particularly preferred embodiment, the isolated oligonucleotide is 100 % homologous with the nucleotide sequence residing between position 555 and position 719 inclusive of Sequence I.D. No. 1. Exemplary high stringency hybridization conditions are provided in the Examples.

Alternatively, the isolated oligonucleotide is capable of hybridizing under stringent conditions to a unique fragment of the nucleotide sequence residing between positions 555 and 719 of Sequence I.D. No.1. As used herein, the phrase "unique fragment" refers to a nucleic acid sequence having less than 25% sequence homology with previously identified nucleic acid sequences. More preferably, the unique fragments have less than 10% sequence homology with known nucleic acid sequences. Such unique fragments can be identified by searching the Genbank, PIR and/or Swiss-Prot data bases (e.g., release date January 20, 1994) using the Eugene program available through the Harvard Molecular Biology Core Research Resource, Cambridge, MA. For example, using the entire X domain, several classes of intranuclear DNA binding proteins and intracytoplasmic proteins were identified as having some degree of sequence homology with the X domain. These proteins included neurofilament triplet M protein, nucleolin, troponin T, alphaglobulin type B precursor (tandem 1), legumin B LegJ precursor, major centromere autoantigen CENP-B, calreticulin precursor (clone 3) and non-histone chromosomal protein HMG-1.

These proteins had, at most, about 50% sequence homology over 35 amino acids or about 35% sequence homology over 54 amino acids. The regions of sequence homology were located in the negatively charged, glutamate-rich region, with less than 20% sequence homology observed for the relatively uncharged portion of the X domain (Sequence I.D. No. 6).

In view of the many DNA binding proteins which contain highly negatively charged regions for binding to DNA, it is likely that the observed sequence homology between the highly charged region of the X domain and the above-mentioned intranuclear DNA binding proteins is coincidental and does not reflect a similar functional activity (e.g., the ability to bind to a ligand expressed on an epithelial cell) between the X domain and these DNA binding proteins. Hence, Sequence I.D. Nos. 6 and 7 are deemed to be unique fragments of the  $\alpha^E$  subunit amino acid and nucleotide sequences, respectively. Accordingly, in a particularly preferred embodiment, the isolated oligonucleotide is Sequence I.D. No. 7 (the sequence corresponding to nucleotides 555 through 656 inclusive of the  $\alpha^E$  cDNA). The unique fragments are useful, for example, as probes and primers in nucleic acid hybridization assays and in amplification reactions, respectively.

For applications directed to the use of an isolated oligonucleotide for regulating transcription and/or translation of the  $\alpha^E$  subunit, the preferred oligonucleotide is an antisense oligonucleotide between about 10 and about 100 nucleotides in length. The antisense oligonucleotide is capable of hybridizing under high stringency conditions to the above-described unique fragments of Sequence I.D. No. 1. As used herein, "antisense oligonucleotide" refers to an oligonucleotide (DNA and/or RNA) that is capable of hybridizing to the naturally-occurring DNA or mRNA encoding the novel  $\alpha^E$  subunit of human integrin. In a preferred embodiment, the antisense oligonucleotide is capable of hybridizing in vivo to the nucleotide sequence residing between positions 555 and 656 inclusive of Sequence I.D. No. 1 or its transcription product. Base-pairing of the antisense oligonucleotide with the DNA (or RNA) encoding the uncharged region of the  $\alpha^E$  subunit in vivo, prevents localization of intra-epithelial lymphocytes by preventing transcription (or translation) of the  $\alpha^E$  subunit.

Methods for expressing the above-identified oligonucleotides

in a suitable expression system including a host cell are well known to those of ordinary skill in the art (see, e.g., Sambrook, et al., Molecular Cloning, A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press, Plainview, NY (1989)). The term "host cell" refers to a prokaryotic or eukaryotic cell which, together with a recombinant vector, comprises an expression system. The term host cell also embraces a host cell in which the vector or isolated oligonucleotide has integrated into the host cell nucleic acid. In a preferred embodiment, the expression vector includes at least one strand of the above-disclosed isolated oligonucleotide. Preferably, the oligonucleotide is operatively joined to at least one regulatory sequence, e.g., a promoter sequence, an enhancer sequence. A coding sequence (e.g., the isolated oligonucleotide) and a regulatory sequence are said to be operably joined when they are linked in such a way as to place expression of the coding sequence under the influence or control of the regulatory sequence.

Suitable cell lines include mammalian cells (e.g., Chinese hamster ovary cells (CHO), monkey COS-1 cell); bacterial cells (e.g., E. coli, B. subtilis and Pseudomonas strains); insect cells (e.g., SF9) and various yeast strains. Exemplary procedures for obtaining expression of a foreign gene in the above-identified cell lines are disclosed in U.S. 5,211,657, the entire contents of which are incorporated herein by reference.

### Examples

The methodology for construction, isolation and characterization of the integrin  $\alpha^E$  cDNA clone of the present invention is described in detail in Example 1. An exemplary method for identifying peptide analogs which inhibit  $\alpha^E\beta_7$ -mediated adhesion between intra-epithelial T lymphocytes and epithelial cells is described in detail in Example 2. All references recited in this application are incorporated in their entirety herein by reference.

#### Example 1. Cloning and Characterizing the $\alpha^E$ subunit cDNA.

##### MATERIALS AND METHODS

**Purification and sequencing of the HML-1  $\alpha$  chain.** Purification of the  $\alpha^E\beta_7$  complex was performed by immunoaffinity column chromatography from a human hairy cell leukemia spleen lysate using the Ber-ACT8 monoclonal antibody, as previously described (Parker, C.M. et al., (1992) Proc. Natl. Acad. Sci. U.S.A. 89, 1924-1928). The HML-1 antigen was then resolved by SDS-PAGE under non-reducing conditions (Hochstenback, F. et al., (1988) J. Exp. Med. 168, 761-776), and the region of the gel containing the larger 175 kDa subunit was excised and equilibrated in reducing buffer: 125 mM Tris-Cl, 0.1% (w/v) SDS, 5% (v/v) 2-mercaptoethanol, and 10% (v/v) glycerol. The two resulting  $\alpha^E$  fragments of 150 kDa and 25 kDa were then separated on a second SDS/PAGE gel under reducing conditions and electroblotted to a poly(vinylidene difluoride) membrane (Problott: Applied Biosystems) in 10 mM CAPS pH 11, 10% Methanol. The HML-1  $\alpha$  subunit 150 kDa and 25 kDa fragments were visualized by Coomassie blue staining and sequenced using an Applied Biosystems model 470A gas-phase sequencer equipped with a model 120A phenylthiohydantoin amino-acid sequencer (Harvard Microsequencing Facility, Cambridge, MA). Approximately 9 pMol of the 150 kDa fragment and 19 pMol of the 25 kDa fragment were analyzed.

**Preparation of cDNA library.** RNA was prepared using the urea/LiCl method (Auffray, C. et al., (1980) Eur. J. Biochem 107, 303-314) from  $3 \times 10^8$  IEL that had been cultured for 5 days with 2 ng/ml TGF- $\beta$ 1 (gift of Celtrix, Palo Alto, CA). Poly (A) RNA was selected and cDNA was synthesized from this RNA in the presence of methylmercuric hydroxide using both random oligonucleotides as well

selectively remove low molecular weight species (under 500 bp) and the two types of cDNA were mixed, methylated, and ligated into the lambda zapII vector (Stratagene, La Jolla, CA). The library contained inserts ranging from 0.6 - 5 kb, with  $1.4 \times 10^6$  independent clones (Clontech laboratories Inc., Palo Alto, CA).

**Isolation of cDNA clones.** Based upon the determined N-terminal amino acid sequences of the  $\alpha^2$  150 kDa and 25 kDa fragments, two degenerate oligonucleotide probes were generated (Fig. 1). Inosine residues were used to reduce the degree of degeneracy. The amplified cDNA library was screened by plaque filter hybridization in 6X SSC(27), 0.5% (w/v) SDS, 50 mM HEPES, pH 7, 5X Denhardt's solution(27), and 1 mM EDTA, with each of these  $^{32}$ P end-labelled oligonucleotides using standard techniques (Sambrook, J. et al., (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press). Filters were washed at low stringency at room temperature in 2X SSC, 0.5% (w/v) SDS. The filters were then washed under stringent conditions, at 37°C for 10 minutes (oligonucleotide E: 2X SSC, 0.5% (w/v) SDS; oligonucleotide B: 3 M tetramethylammonium chloride (Aldrich), 2 mM EDTA, 50 mM tris-Cl pH 8). Plaques which hybridized with both oligonucleotides were purified. Each subsequent screening was performed with cDNA restriction fragments isolated from clones identified in earlier screenings. These restriction fragments were labelled using random-priming hexamers, and the filters were hybridized under stringent conditions in 4.8 X SSC, 50 mM HEPES, pH 7, 0.5%(w/v) SDS, 5X Denhardt's solution, 50% formamide, and 200 ug/ml sheared denatured salmon sperm DNA (Sigma). Filters were stringently washed in 0.2% SSC, 0.1% SDS, at 37°C for 1 hour. Positive clones were subcloned in the Bluescript SK(-) plasmid (Stratagene), with the helper phage R407, using an in vivo excision protocol (Short, J.M. et al., (1988) Nucleic Acids Res. 16, 7583-7600), and restriction mapped using standard procedures (Sambrook, J. et al., (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press).

**Nucleotide sequencing.** The ends of inserts of positive clones were sequenced using the T3 and T7 primer regions from the Bluescript SK(-) vector. Additionally, a series of deletion clones were generated from selected inserts using Exonuclease III (Erase-a-base kit, Promega) in order to sequence their interior regions.



Sequencing reactions were performed using the double-standard dideoxy-termination method (Sequenase kit, USBiochem) and the sequence was analyzed by computer using DNASTar, DNA Strider, and the Genetics Computer Group software package (Madison, WI) on the Molecular Biology Computer Research Resource Facility (Dana Farber Cancer Institute, Boston, MA). Many clones isolated had extraneous sequence attached to one or both ends (Fig. 1), which was not present in any other clone. However, it was possible to recognize the  $\alpha^E$  encoding sequence based on overlaps with other previously identified clones, i.e., those clones identified in a prior round of screening. For example, clone 38 (identified in a first round of screening), was used to identify clone 1-39A. A 3' fragment of clone 1-39A then was used to rescreen the library and identify clone 2-54. Thereafter, a 3' fragment of clone 2-54 was used to identify clone 3-15.

**Monoclonal antibodies and cell lines.** Previously described mAb used were HML-1 (mouse anti-human  $\alpha^E\beta$ , IgG2a) (Cerf-Bensussan, N. et al., (1987) Eur. J. Immunol. 17, 1279-1285), Ber-ACT8 (mouse anti-human  $\alpha^E\beta$ , IgG1) (Kruschwitz, M. et al., (1991) J. Clin. Path 44, 636-645), OKT3 (mouse anti-human CD3, IgG2a, available from American Type Culture Collection (ATCC), Bethesda, MD), OKT4 (mouse anti-human CD4, IgG2b, ATCC), OKT8 (mouse anti-human CD8\*, ATCC), 64.1 (mouse anti-human CD3) (Hansen, J.A. et al., (1984) in Leukocyte typing: Human Leukocyte Differentiation Antigens detected by Monoclonal antibodies, 195-212), SPV-T3b (mouse anti-human CD3, IgG2a) (Spits, H. et al., (1983) Hybridoma 2, 423-437), 3C10 (mouse anti-human CD14, IgG2b) (Van Voorhis, W.C. et al., (1983) J. Exp. Med. 158, 126-145), HB12B (mouse-anti human CD19, IgG1) (Kansas, G.S. et al., (1991) J. Immunol. 147, 4094-4102), B1 (mouse anti-human CD20, IgG2a) (Stashenko, P. et al., (1980) J. Immunol. 125, 1678-1685) and the anti- $\beta$ , C-terminal peptide (amino acids 736-755) antiserum (Parker, C.M. et al., (1992) Proc. Natl. Acad. Sci. U.S.A. 89, 1924-1928).

To produce an anti- $\alpha^E$  C-terminal peptide serum, a peptide was synthesized corresponding to residues 1136-1160 deduced to lie in the  $\alpha^E$  cytoplasmic tail, with an added cysteine residue at the amino terminus of the peptide to permit coupling to carrier protein. The peptide was conjugated to keyhole limpet hemocyanin (Pierce, Rockford, IL), using the heterobifunctional cross linking

agent,  $\alpha$ -Maleimidobenzoyl-N-hydroxysulfosuccinimide (Pierce), according to the manufacturer's instructions. Rabbit antisera were generated by immunization with the conjugated  $\alpha^E$ 1136-1160 peptide using standard protocols (Harlow, E. et al., (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press).

To produce an anti- $\alpha^E$  X domain (charged fragment) peptide serum, a peptide was synthesized containing ALEKEEEEDKEEEDEEEEC (i.e., Sequence I.D. No. 5 with an added cysteine residue at the carboxy-terminus of the peptide to permit coupling to carrier protein) and antibodies were generated using the same protocol as described above for generating peptides to the  $\alpha^E$  C-terminal peptide. The antiserum produced thereby was capable of immunoprecipitating the  $\alpha^E\beta$ , heterodimer.

iIEL lines were derived and cultured as previously described (Cepek, K.L. et al., (1993) J. Immunol. 150, 3459-3470). Cultured tumor lines (PEER, MOLT 4, MOLT 13, MOLT 17, HPB-ALL, HUT 78, JY', VA2, RD, HeLa, A431) were maintained in RPMI-1640 (GIBCO, Grand Island, NY) containing 10% fetal bovine serum (Hyclone, Logan, UT). The epithelial cell line 16E6.A5 was maintained in DFCI-1 medium (Cepek, K.L. et al., (1993) J. Immunol. 150, 3459-3470).

Leukocyte concentrates produced as a by-product from plateletphoresis of randomly selected healthy donors were used as a starting material to isolate fresh peripheral blood monocytes and T cells. Peripheral blood mononuclear cells (PBMC) were separated by density gradient centrifugation (Ficoll-Plaque, Pharmacia LKB, Piscataway, NJ) and monocytes were isolated by adherence to polystyrene tissue culture flasks (Porcelli, S. et al., (1992) Nature 360, 593-597) and used to prepare RNA (see below). A fraction of the adherent cells in a separate flask were detached by incubation at 37°C in PBS/0.5 mM EDTA, and analyzed by flow cytometry using the FACScan flow cytometer (Beckton Dickinson, Mountain View, CA) to reveal 91% CD14+, 2% CD19/20+, and <1% CD3+ cells. T lymphocytes were isolated from the non-adherent PBMC using an anti-CD3 antibody, 64.1, and positive selection by magnetic bead separation (Dynabeads M-450, Dynal A.S., Oslo, Norway) according to the manufacturer's instructions. To isolate B lymphocytes, surgically removed fresh tonsils were teased into fragments with forceps and forced through a wire mesh. The resulting cell suspension was subjected to density gradient

centrifugation (Ficoll-Paque) to obtain mononuclear cells. Tonsillar B cells were further purified by depleting T lymphocytes with mAbs recognizing CD3 (64.1), CD4 (OKT4) and CD8 (OKT8) and negative selection by magnetic bead selection. An aliquot of the negatively selected population was analyzed by flow cytometry: 97% were positive with anti-CD19 + anti-CD20 mAbs (HB-12b and B1) and 1.3% were positive with anti-CD3 (SPV-T3b).

Neutrophils were purified from peripheral blood as previously described (Neuman, E. et al., (1992) J. Immunol. 148, 3520-3527). Differential analysis was performed using Wright's/Giemsa stained cytopsin preparations, and revealed 92% neutrophils, with 8% eosinophils.

**Immunoprecipitations.** TGF- $\beta$ 1 treated cultured iEL (35 x 10<sup>6</sup> cells) were surface radioiodinated with 1 mCi of Na<sup>125</sup>I using lactoperoxidase, lysed in 1 ml of 0.5% Triton-X100, 8mM iodoacetamide, 10 mM phenylmethylsulfonyl fluoride in TBS (50 mM tris-Cl, pH 7.6, 140 mM NaCl), and precleared with 4 ul normal rabbit serum, followed by 200 ul of Staphylococcal A Cowan I. Following centrifugation, supernatant containing 3 x 10<sup>6</sup> cell equivalents was used for each specific immunoprecipitation, using 5 ul of antiserum, or 0.1 ul of HML-1 mAb ascites, as previously described (Hochstenbach, F. et al., (1988) J. Exp. Med. 168, 761-776). Following a 1 hour incubation at 4°C, the supernatant was incubated for 1 hour with 75 ul of protein A-sepharose (10%, w/v, in TBS). The immune complexes were washed five times in wash buffer (0.1% v/v Triton-X100/TBS), eluted with sample buffer (5% glycerol, 1.5% (w/v) SDS, 0.2% (w/v) bromophenol blue, 188 mM tris-Cl pH 8.8) and resolved on a 7% denaturing polyacrylamide gel under non-reducing conditions (Hochstenbach, F. et al., (1988) J. Exp. Med. 168, 761-776).

In the reprecipitation experiment, 7 X 10<sup>7</sup> cultured iEL were radioiodinated with 2.5 mCi, lysed and precleared as above. A specific immunoprecipitation was performed with the entire lysate, using 2 ul of HML-1 mAb ascites, followed by 300 ul of protein A-sepharose. After 5 washes in wash buffer,  $\alpha^B$ , was eluted from the matrix by heating in a boiling water bath for 5 minutes in 150 ul of 1% (w/v) SDS. Following brief centrifugation, the supernatant was collected and treated with 150 ul precleared fetal calf serum for 30 minutes. Four ml of 2% (v/v) Triton-X100/TBS

were added, and the supernatant incubated overnight at 4°C. Equal aliquots of 600 ul were then precleared and immunoprecipitated with the anti- $\alpha^E\beta$ , mAb HML-1 (0.1 ul), the anti- $\alpha^E$ 1136-1160 C-terminal peptide serum (10 ul), followed by protein A-sepharose (75 ul). Immune complexes were washed and the reprecipitated proteins eluted, and resolved by SDS-PAGE under non-reducing conditions, as previously described (Hochstenbach, F. et al., (1988) J. Exp. Med. 168, 761-776).

**Northern Blot Analysis.** RNA was prepared from cell lines and leukocyte fractions using guanidium isothiocyanate lysis, followed by acid-phenol/chloroform extraction (Chomczynski, P. et al., (1987) Anal. Biochem 162, 156-159). Poly(A) RNA was isolated from total RNA using standard techniques (Sambrook, J. et al., (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press). RNA was resolved on 1.2% agarose-formaldehyde gels and transferred to nylon membranes (Hybond N: Amersham). Commercially available Northern blots (Clontech) were used to analyze human tissue RNA. Blots were hybridized at 42°C in 5 X SSPE (Sambrook, J. et al., (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press), 10 X Denhardt's solution, 50% (v/v) formamide, 2% (w/v) SDS, and 100 ug/ml salmon sperm DNA. Blots were washed in 0.1 X SSC, 0.1% (w/v) SDS, at 50°C, for one hour.

The probes used to analyze the Northern blots included the  $\alpha^E$  cDNA probe which consisted of two restriction fragments in approximately equimolar amounts and spanning approximately 1/3 of the deduced coding sequence, extending from nucleotide 1915-2394 (1-39A Apa I-BamH I) and 2394-3084 (1-39A BamH I-EcoR I). The  $\beta$  probe extended from nucleotide 47-2801 of the published sequence (Yuan, Q.A. et al., (1990) Int. Immunol. 2, 1097-1108). The glyceraldehyde 3-phosphate dehydrogenase (GAP) probe (Clontech) extended from nucleotide 71-1053 of the published sequence (Arcari, P. et al., (1984) Nucleic Acids Res. 12, 9179-9189).

## RESULTS

### Purification and N-terminal amino acid sequence of the $\alpha^E$ chain.

To determine the primary amino acid sequence of the  $\alpha^E$  protein, the HML-1 ( $\alpha^E\beta_3$ ) complex was purified from a hairy cell leukemia spleen lysate by immunoaffinity column chromatography, followed by two stages of SDS-PAGE (see Materials and Methods). The  $\alpha^E$  subunit resolved under reducing conditions into 150 kDa and 25 kDa fragments. A sequence of 20 amino acids was determined from the N-terminus of the 25 kDa fragment (Fig. 1). Surprisingly, it was this sequence, not that from the 150 kDa fragment, which bore homology to known integrin  $\alpha$  chain amino termini. This suggested that  $\alpha^E$  was cleaved towards the N-terminal portion of the precursor polypeptide, unlike other integrins which are cleaved proximal to the C-terminus or near the midpoint. The N-terminal sequence from the 150 kDa fragment showed a preponderance of negatively charged residues (Fig. 1), with no significant homology with previously identified integrin  $\alpha$  chains.

**Cloning of a cDNA encoding  $\alpha^E$ .** To clone a cDNA encoding  $\alpha^E$ , a lambda Zap II cDNA library was synthesized from TGF- $\beta$ 1 treated iIEL mRNA using both oligo d(T) and random priming. Based on the amino acid sequence obtained from the 150 and 25 kDa fragments, corresponding degenerate oligonucleotides were synthesized, one from each, and were used to screen the library (Fig. 1). Analysis of 420,000 phage plaques resulted in the identification of one cDNA clone, 38, that hybridized with both oligonucleotides. This clone was purified and the insert cDNA was sequenced. Both of the determined amino acid sequences were encoded by clone 38 (see below), indicating that this clone was derived from authentic  $\alpha^E$  mRNA. The entire 38 insert (0.8) was used to rescreen the library to identify clones that extended further in the 3' direction. Probes derived from the 3' ends of these and of subsequently identified cDNA clones were in turn used to successively rescreen the library. All clones used to produce the composite sequence were shown to detect transcripts of the same size in Northern blot analysis of TGF- $\beta$ 1 treated iIEL RNA. After five rounds of screening, a composite sequence was determined as the  $\alpha^E$  coding sequence from cDNA clones 38, 1-39A, 2-54, and 3-15 (Fig. 1).

Extraneous sequences on the ends of the individual clones (broken lines) and sequences determined from insert ends or from

deletion clones (horizontal arrows) also are illustrated in Fig. 1. In regions where the nucleotide sequence encoding  $\alpha^E$  differed between clones, short stretches of other cDNA clones were used to establish a consensus. Two of these changes did not result in a frameshift in the open reading frame and may stem from genuine polymorphisms amongst alleles of  $\alpha^E$ . Alternatively, differences may represent artifactual point mutations in the cDNA clones generated during synthesis of the cDNA library or in its subsequent propagation. Specifically, at nucleotide 2000, cytosine was replaced by adenosine in about one-third of the clones, resulting in a change in amino acid 341 from aspartic acid (D) to glutamic acid (E). At nucleotide 2190, cytosine was deleted in approximately one-sixth of the clones, resulting in a frameshift. At nucleotide 3242, guanosine was replaced by cytosine in about one-fifth of the clones, resulting in a change in amino acid 1022 from cysteine (C) to serine (S).

$\alpha^E$  Sequence Analysis. A total of 3.9 kb of composite cDNA sequence was determined, which contained a 3.5 kb open reading frame (see Sequence I.D. Nos. 1 and 2). The 125 bp 5' untranslated region ended with a sequence characteristic of a translation initiation site (Hogervorst, F. et al., (1991) Eur. J. Biochem, 199, 425-433) and a methionine codon (Sequence I.D. No. 1). This presumed initiation codon was the first residue in a stretch of 18 hydrophobic amino acids that was inferred to encode a signal sequence. This sequence ended in accordance with the (-3, -1) rule of von Heijne (von Heijne, G. (1986) Nucleic Acids Res. 14, 4683-4690), suggesting a likely site for cleavage by signal peptidase. The next 20 deduced amino acids were identical to those determined from peptide sequencing of the 25 kDa fragment. From amino acids 160-169 (Sequence I.D. No. 1), a deduced sequence identical to the determined peptide sequence from the 150 kDa fragment was found, confirming the smaller fragment to be N-terminal. Just prior to the second (150 kDa) peptide site were a pair of basic residues (RR). Similar motifs are found in several other integrin  $\alpha$  subunits in which the cleavage site is preceded by two basic residues (Tamura, R.N. et al., (1990) J. Cell Biol. 111, 1593-1604; Teixido, J. et al., (1992) J. Biol. Chem. 267, 1786-1791; Takada, Y. et al., (1991) J. Cell. Biol. 115, 257-266; Argraves, W. et al., (1987) J. Cell. Biol. 105, 1183-1190; Song, W.K. et al., (1992) J.

Cell. Biol. 117, 643-657; Bossy B. et al., (1991) EMBO J. 10, 2375-2385; Suzuki, S. et al., (1987) J. Biol. Chem. 262, 14080-14085; Poncz, M. et al., (1987) J. Biol. Chem. 262, 8476-8482), and may be the recognition site of an eukaryotic subtilisin-like protease (Barr, P.J. (1991) Cell 66, 1-3). A single potential N-glycan acceptor site was present in the first 159 residues which comprised the smaller fragment.

The deduced N terminus of the larger, 150 kDa, fragment was characterized by a highly charged sequence (Sequence I.D. No. 1, amino acids 160-179). Following the first two amino acids of this fragment (AL), 16 out of the following 18 amino acids were acidic and the remaining two were basic. Within the 150 kDa fragment deduced amino acid sequence, nine potential N-linked glycosylation sites were found, making a total of ten in the entire  $\alpha^E$  sequence. A region of 23 hydrophobic amino acids (amino acids 1106-1128, Sequence I.D. No. 1) was predicted to encode a transmembrane region, followed by a cytoplasmic tail of 32 residues and a stop codon. Thus, the composite  $\alpha^E$  cDNA predicted a typical type I transmembrane protein that was post-translationally cleaved towards the N terminus. The 270 bp of the predicted 3' untranslated sequence included a consensus polyadenylation signal (Sequence I.D. No. 1, aataaa) and a poly(A) tail.

The deduced amino acid sequence predicted a small fragment with a size of 20 kDa, allowing 2.5 kDa for its single potential N-linked glycosylation site. This value approximated the observed migration in SDS-PAGE of the smaller  $\alpha^E$  subunit (25 kDa). With nine N-linked glycosylation sites, the predicted molecular mass for the larger fragment was 133 kDa, somewhat smaller than the estimate from SDS-PAGE of 150 kDa. This discrepancy between predicted molecular mass and migration in SDS-PAGE is similar to that observed for other integrin  $\alpha$  subunits (Tamura, R.N. et al., (1990) J. Cell. Biol. 111, 1593-1604; Argraves, W. et al., (1987) J. Cell. Biol. 105, 1183-1190; Ignatius, M.J. et al., (1990) J. Cell. Biol. 111, 709-720). The predicted pI of this larger subunit (5.2) may account for the difference in predicted and determined molecular weights as some negatively charged proteins migrate more slowly in SDS-PAGE than predicted from deduced molecular mass (Georges, E. et al., (1987) Eur. J. Biochem. 165, 281-287; Matagne, A. et al., (1991) Biochem. J. 280, 553-556).

### Comparison of $\alpha^E$ with other integrin $\alpha$ subunits.

The deduced amino acid sequence of  $\alpha^E$  was manually aligned with that of other integrin  $\alpha$  chains. The locations of cysteine and proline residues are often important in secondary structure of proteins, and were used as a primary basis for alignment, as were regions where blocks of amino acids showed strong homology with other integrin  $\alpha$  subunits. In regions where homology was not readily apparent, gaps were arbitrarily introduced in the sequences in order to maximize alignment.

Comparison of the deduced  $\alpha^E$  amino acid sequence (Sequence I.D. No. 2) with that of other integrin  $\alpha$  subunits revealed the presence of many conserved features. Fourteen out of a total of 25 cysteine residues were conserved between  $\alpha^E$  and other  $\alpha$  chains. Immediately following the highly charged regions, a region of 193 amino acids homologous to typical I domains was present (Sequence I.D. No. 2, amino acids 181-372). Like other integrin  $\alpha$  subunits, the  $\alpha^E$  cDNA encoded seven repeated regions, the last three of which were potential divalent cation binding sites, resembling the EF-hand loop motif, with a consensus sequence corresponding to DXXXDXXXD (Tuckwell, D.S. et al., (1992) Biochem J. 285, 325-331). The probable intracytoplasmic region contained the highly conserved GFFKR sequence (Sequence I.D. No. 1), which is present in every sequenced integrin  $\alpha$  chain except the Drosophila PS2  $\alpha$  chain (GFFNR) (Brown, N.H. et al., (1989) Cell 59, 185-195), the chicken  $\alpha^E$  chain (GFFDR) (Bossy, B. et al., (1991) EMBO J. 10, 2375-2385), and the hamster  $\alpha^{3B}$  chain (DFFKP) (Tamura, R.N. et al., (1991) Proc. Natl. Acad. Sci. U.S.A. 88, 10183-10187).

Immediately preceding the presumed I domain was a stretch of 55 amino acids which did not align with other known  $\alpha$  chains (Sequence I.D. No. 2, amino acids 126-180). For descriptive purposes, this region will be referred to as the 'extra' or X domain. Interestingly, the highly charged region described above (see  $\alpha^E$  sequence analysis) and the unusually located cleavage site are within this X domain.

A homology tree was generated with the deduced amino acid sequence of  $\alpha^E$  as compared to those of other integrin  $\alpha$  subunits, using the computer program PILEUP. The program was executed both with complete sequences, as well as the sequences from which all I domains, and the X domain of  $\alpha^E$  were removed. This was done to



facilitate computer alignment of I domain containing integrin  $\alpha$  subunits with those that lack this region. In each case, the overall shape of the homology tree generated was similar, although the branch point of  $\alpha^E$  was slightly more closely related to the  $\beta_2$  associated  $\alpha$  chains ( $\alpha^L$ ,  $\alpha^M$ ,  $\alpha^X$ ) than the  $\beta_1$  associated  $\alpha$  chains ( $\alpha^1$  and  $\alpha^2$ ) (Figs. 2 and 3).

Confirmation that the cloned cDNA encodes  $\alpha^E$ . A rabbit polyclonal antiserum was generated to a synthetic peptide corresponding to residues 1136-1160 of the deduced  $\alpha^E$  amino acid sequence. This anti- $\alpha^E$  C-terminal peptide antiserum was used to immunoprecipitate polypeptides from a lysate of  $^{125}\text{I}$ -labelled cultured iIEL. When this immunoprecipitate was resolved by SDS-PAGE under non-reducing conditions, a complex of polypeptides was observed, indistinguishable from that immunoprecipitated by the HML-1 mAb. This included the 175 kDa  $\alpha^E$  subunit, the 105 kDa  $\beta_1$  subunit, and two less well visualized intermediate sized polypeptides of 155 and 135 kDa (Fig. 4, lanes 2 and 3).

Next, the HML-1 complex was isolated by immunoprecipitation with the HML-1 mAb followed by elution of the component chains with SDS and heat. Equal aliquots of this immunoprecipitated and chain separated HML-1 complex were then reprecipitated with the HML-1 mAb, the anti- $\alpha^E$  C-terminal peptide antiserum, or an anti- $\beta_1$  C-terminal peptide antiserum (Parker, C.M. et al., (1992) Proc. Natl. Acad. Sci. U.S.A. 89, 1924-1928) and resolved by SDS-PAGE under non-reducing conditions. The anti- $\beta_1$  antiserum reprecipitated the 105 kDa  $\beta_1$  polypeptide, but not the other species (Fig. 4, lane 6). The anti- $\alpha^E$  antiserum reprecipitated the  $\alpha^E$  175 kDa polypeptide as well as the 155 and 135 kDa species, but not the 105 kDa  $\beta_1$  polypeptide (Fig. 4, lane 5). This indicated that the  $\alpha^E$ -antiserum recognized a polypeptide that was not only similar in size to, but also immunochemically cross-reactive with the polypeptides of 175, 155 and 135 kDa, recognized by the HML-1 mAb. Since the anti- $\alpha^E$  antiserum was generated to a deduced C-terminal amino-acid sequence that was distinct from the determined N-terminal amino acid sequences used to isolate the cDNA, these experiments offer independent evidence that the cDNA isolated encodes the same  $\alpha^E$  subunit recognized by the HML-1 antibody. The reprecipitation experiment also offers clues to the identities of the 155 and 135 kDa species. Either these two

additional polypeptides reassociated with the 175 kDa species after boiling in SDS, or, more likely, they directly interacted with the anti- $\alpha^E$  C-terminal peptide antiserum. Thus, these may be alternate structural forms of  $\alpha^E$ . Alternative RNA splicing (Tamura, R.N. et al., (1991) Proc. Natl. Acad. Sci. U.S.A. 88, 10183-10187; Tamura, R.N. et al., (1990) J. Cell. Biol. 111, 1593-1604; Hogervorst, F. et al., (1991) Eur. J. Biochem. 199, 425-433; Bray, P.F. et al., (1990) J. Biol. Chem. 265, 9587-9590; Zimrim, A.B. et al., (1990) J. Biol. Chem. 265, 8590-8595), alternative proteolytic processing (Calvete, J.J. et al., (1990) Febs Letters 272, 37-40; Loftus, J.C. et al., (1988) J. Biol. Chem. 263, 11025-11028), differential glycosylation (Sonnenberg, A. et al., (1990) J. Cell. Sci. 96, 207-217; Kim, L.T. et al., (1992) J. Cell. Sci. 103, 743-753; Hotchin, N.A. et al., (1992) J. Biol. Chem. 267, 14852-14858; Bednarczyk, J.L. et al., (1992) Clin. Exp. Metastasis 10, 281-290) and alternative tertiary structure of the same polypeptide backbone (Teixido, J. et al., (1992) J. Biol. Chem. 267, 1786-1791; Rubio, M. et al., (1992) Eur. J. Immunol. 22, 1099-1102) have all been reported to occur in other integrin  $\alpha$  subunits, resulting in multiple cell-surface forms, and may account for this observation.

**Distribution of  $\alpha^E$  mRNA transcripts.** An  $\alpha^E$  cDNA probe which spanned 1.2 kb of the coding sequence was used to analyze the distribution of  $\alpha^E$  mRNA in normal tissues and in in vitro cultured lines. Freshly isolated human iIEL serve as a model of fresh iIEL in their surface expression of adhesion molecules (Cepek, K.L. et al., (1993) J. Immunol. 150, 3459-3470). Northern blot analysis of RNA from a TGF- $\beta$ 1 treated iIEL line indicated a single species of 4.5 kb that hybridized strongly with the  $\alpha^E$  cDNA probe (Fig. 5A, seen as a broad smear in this exposure). As expected, the  $\alpha^E$  species was also readily detected in RNA from the malignant hairy cell leukemia infiltrated spleen used to isolate protein for N-terminal amino acid sequencing. Moderate to low levels of the  $\alpha^E$  mRNA were also detected in poly (A) enriched RNA from a range of normal human tissues (Fig. 5B) including lung, pancreas, spleen, thymus, prostate, testis, ovary, small intestine, colon, and peripheral blood leukocytes. To determine which leukocytes transcribe  $\alpha^E$  mRNA, freshly isolated leukocytes were fractionated into their component sub-populations and analyzed by Northern analysis. Barely detectable levels of the  $\alpha^E$  mRNA were found in

poly (A) RNA from peripheral blood monocytes and T cells, as well as from tonsillar B cells, but not from neutrophils (Fig. 5C). Note that the tissue samples (Fig. 5B) contained 2-4 fold more RNA than did the cultured iIEL sample (Fig. 5A). Yet the  $\alpha^E$  mRNA in cultured iIEL was much more strongly detected than in any of the normal tissues, indicating that  $\alpha^E$  mRNA levels in these tissues were markedly lower than in iIEL. The  $\alpha^E$  species was not detected in RNA from other tissues, including heart, brain, placenta, liver, skeletal muscle and kidney.

Analysis of tumor and transformed cell lines also was performed. In a panel of T cell tumors both  $\gamma\delta$  T cell receptor bearing cells (PEER and MOLT 13) and  $\alpha\beta$  T cell receptor bearing cells (HPB-ALL and MOLT 17) showed the  $\alpha^E$  4.5 kb mRNA. This transcript was not detected in other T cell tumors, including MOLT 4, HUT 78, nor in JY' (an Epstein-Barr virus-transformed B cell line selected for high surface expression of the  $\beta_2$  protein) (Fig. 5D). Amongst non-lymphocytic transformed or tumor cell lines, the 4.5 kb  $\alpha^E$  mRNA species also was present in VA2 (fibroblast line), RD (rhabdomyosarcoma line), HeLa (cervical carcinoma line, epithelial like), and barely detected in 16E6.A5 but not in A431 (epithelial cell lines) (Fig. 5E).

Interestingly, in RNA from MOLT 13, 2.3 kb and 1.3 kb transcripts were detected, which hybridized more strongly with  $\alpha^E$  cDNA than did the faintly visualized 4.5 kb  $\alpha^E$  transcript (Fig. 5D). Both the 2.3 kb and 1.3 kb species also were detected in PEER, HPB-ALL and MOLT 17 T cell leukemia lines. These transcripts were apparent only upon hybridization with  $\alpha^E$  cDNA, and not with  $\beta_2$  cDNA or the positive control cDNA, suggesting that they were not due to non-specific binding of probe to the filter, but instead may represent cross-hybridization. The 2.3 kb transcript was detected only in RNA from tumor cell lines, but the 1.3 kb species was detected also in thymus and testis (Fig. 5B). Since integrin  $\alpha$  subunit coding sequences are usually > 3.3 kb, it seems unlikely that these transcripts represent alternatively spliced forms of  $\alpha^E$  encoding a traditional integrin structure. These smaller species may be due to coincidental hybridization of a portion of the  $\alpha^E$  cDNA probe to an unrelated transcript.

These same Northern blots were probed with a cDNA spanning the full length coding sequence of  $\beta_2$  in order to correlate the

potential co-expression of  $\beta_7$  with  $\alpha^E$ . The  $\beta_7$  probe detected a 3 kb mRNA that hybridized abundantly with RNA derived from cultured iIEL and hairy cell spleen (Fig. 5A). Multiple tissues and cell lines which showed the presence of the 4.5 kb  $\alpha^E$  transcript also showed the 3 kb  $\beta_7$  transcript, including lung, spleen, thymus, prostate, ovary, small intestine, colon, and peripheral blood leukocytes (Fig. 5B), fresh peripheral blood monocytes and T cells, tonsillar B cells (Fig. 5C), PEER, and also in MOLT 13 and MOLT 17 at barely detectable levels (Fig. 5D). This suggested that often the two transcripts were expressed coordinately. Many cells which transcribed both  $\alpha^E$  and  $\beta_7$  also expressed the cell surface heterodimer as detected by immunostaining with anti- $\alpha^E\beta_7$  mAbs of cultured iIEL (Cepek, K.L. et al., (1993) J. Immunol. 150, 3459-3470), hairy cell spleen (Moller, P. et al., (1990) Amer. J. Path. 136, 509-512; Visser, L. et al., (1989) Blood 74, 320-325), < 2% of fresh peripheral blood lymphocytes (Cerf-Bensussan, N. et al., (1987) Eur. J. Immunol. 17, 1279-1285), tonsil, PEER, MOLT 13, and HPB-ALL<sup>b</sup>. As expected, RNA from some sources showed the presence of the 3.0 kb  $\beta_7$  transcript in the absence of detectable  $\alpha^E$ , including fresh peripheral blood neutrophils (Fig. 5C), HUT 78 and JY' (Fig. 5D). Consistent with these findings, the  $\beta_7$  protein is known to be expressed in association with  $\alpha^4$  and not with  $\alpha^E$ , on the surface of JY' cells (Chan, B.M. et al., (1992) J. Biol. Chem. 267, 8366-8370). Surprisingly,  $\alpha^E$  transcripts were detected in the absence of detectable levels of  $\beta_7$  in RNA samples derived from pancreas and testis (Fig. 5B), HPB-ALL (Fig. 5D), and Va2, RD and HeLa and 16E6.A5 (Fig. 5E). In some tissues, the  $\beta_7$  cDNA probe hybridized with mRNA species different in size from the primary 3 kb message including a 4.2 kb species in hairy cell spleen (Fig. 5A), skeletal muscle, normal spleen, thymus, small intestine, peripheral blood leukocytes (Fig. 5B), and tonsillar B cells (Fig. 5C); a 2.4 kb species in heart, liver, and kidney (Fig. 5B); and a 7 kb species in skeletal muscle. Although alternatively spliced forms of  $\beta_7$  have been reported (Erle, D.J. et al., (1991) J. Biol. Chem. 266, 11009-11016; Yuan Q. et al., (1992) J. Biol. Chem. 267, 7352-7358), the nature of the secondary  $\beta_7$  species reported here have not been evaluated.

## DISCUSSION

The cDNA encoding a novel integrin  $\alpha$  chain,  $\alpha^E$ , was cloned and sequenced revealing a type I transmembrane protein. Several lines of evidence indicated that the deduced amino acid sequence of the cDNA described here encodes the 175 kDa  $\alpha$  subunit of the HML-1 antigen. Both peptide sequences determined from purification of the  $\alpha$  subunit protein were encoded by the determined nucleotide sequence of the cloned gene. In addition, the  $\alpha^E$  C-terminal peptide antiserum immunoprecipitated the same species as did the HML-1 antibody. Further, various biochemical features including a proteolytic cleavage site and the presence of N-linked glycosylation were consistent with known features of the HML-1  $\alpha$  chain.

Like other integrin  $\alpha$  subunits, the deduced amino acid sequence of  $\alpha^E$  contained seven repeated domains (the last three of which included potential divalent cation binding sites), an I domain, a pattern of conserved cysteine residues at positions similar to those in other integrin  $\alpha$  subunits, and a GFFKR amino acid motif in the cytoplasmic domain. The  $\alpha^E$  amino acid sequence was most closely related in overall structural features to the other I domain containing integrins since it contained a typical I domain, had three cation binding site consensus motifs, and lacked a membrane proximal cleavage site. In addition, in homology analysis, the  $\alpha^E$  amino acid sequence was more homologous to the I domain containing  $\alpha$  chains even in regions outside the I domain. However,  $\alpha^E$  was a relatively distant member of the I domain containing group. This implies either that  $\alpha^E$  diverged early from any potential I domain/integrin ancestral gene, or that this gene has undergone an accelerated rate of mutation.

While resembling other integrin  $\alpha$  subunits in overall amino acid sequence, one region within  $\alpha^E$  was unique. This region of 55 amino acids located just N-terminal to the I domain, did not bear homology to other integrin  $\alpha$  chains. The proteolytic cleavage site of the  $\alpha^E$  polypeptide was located between amino acids 159 and 160 in this extra or X domain. The  $\alpha^E$  X domain thus accounted for both the unexpected cleavage of an I domain-containing integrin, as well as the unusual site of cleavage. Immediately following the cleavage site within the X domain was a stretch of 18 consecutive charged residues that comprised the N-terminal portion of the 150

kDa  $\alpha^E$  fragment. This negatively charged stretch of amino acids is likely to be either solvent exposed or complexed to a moiety of the opposite charge. The charged stretch of  $\alpha^E$  may be necessary to make the region accessible to the enzyme responsible for cleavage. Such a unique charged sequence seems likely to play an important role in both the structure and function of  $\alpha^E$ .

Analysis of the genomic structure of the classically cleaved integrin,  $\alpha^{ITb}$  (Heidenreich, R. et al., (1990) *Biochem* 29, 1232-1244), has shown that the residues coding for its cleavage site lie within a single exon, not present in the traditional I domain containing integrins,  $\alpha^M$  (Fleming, J.C. et al., (1993) *J. Immunol.* 150, 480-490) and  $\alpha^X$  (Corbi, A.L. et al., (1990) *J. Biol. Chem.* 265, 2782-2788). This cleavage site-containing exon is thought to be the result of an insertion into an ancestral integrin gene. We speculate that in  $\alpha^E$ , the X domain containing the cleavage site may likewise correspond to an exon inserted just prior to the four exons that comprise the I domain in other integrins.

Previous studies utilizing immunohistochemistry have shown that the HML-1 protein is expressed in a highly restricted manner, primarily on mucosal lymphocytes (Cerf-Bensussan, N. et al., (1987) *Eur. J. Immunol.* 17, 1279-1285). In this study we confirmed the presence of high levels of  $\alpha^E$  and  $\beta$ , mRNA in cultured iIEL. However, when poly(A) RNA from non-mucosal leukocytes was analyzed, barely detectable levels of  $\alpha^E$  mRNA were found in fresh peripheral blood monocytes, T cells, and tonsillar B cells. As activated T and B lymphocytes and macrophages have been reported to express  $\alpha^E\beta$ , on the cell surface (Kruschwitz, M. et al., (1991) *J. Clin. Path* 44, 636-645; Pallesen, G. et al., (1990) *Lancet* 335, 537; Visser, L. et al., (1990) *Brit. J. Haematol.* 75, 359-365) it is possible that the  $\alpha^E$  mRNA detected represents the small fraction amongst these predominantly resting cells that are activated. It is also possible that these low quantities are indicative of a basal level of transcription that may be upregulated in response to unknown stimuli, potentially conferring adhesive properties to these leukocytes mediated by  $\alpha^E\beta$ . When poly(A) RNA from a panel of tissues was analyzed, both  $\alpha^E$  and  $\beta$ , mRNA also were detected in tissues known to have IEL, such as lung, small intestine and colon, and in lymphoid tissues, such as thymus and spleen.  $\alpha^E$  and  $\beta$ , mRNA also were present at low levels in non-lymphoid tissues such as

prostate and ovary, and  $\alpha^E$  was observed in the absence of detectable  $\beta$ , in pancreas and testis. It is not known whether the  $\alpha^E$  mRNA is present in tissue parenchymal cells, or is expressed in tissue resident leukocytes. Overall, the distribution of high levels of  $\alpha^E$  mRNA supports the immunohistochemistry data in suggesting that cell-surface  $\alpha^E\beta$ , expression occurs predominantly on mucosal lymphocytes such as IEL and lamina propria lymphocytes in vivo. This restricted distribution implies that  $\alpha^E\beta$ , is important in localization of lymphocytes to mucosal tissues and/or a site-specific lymphocyte function.

Identification of the  $\alpha^E\beta$ , integrin and the genes that encode it make it possible to identify individuals with  $\alpha^E$  or  $\beta$ , deficiency or produce targeted gene knock-out mice. The phenotype of these individuals may serve to illuminate additional functions of the  $\alpha^E\beta$ , molecule in vivo.

Example 2. An adhesion assay for selecting functionally equivalent peptide analogs.

The adhesion assay described herein is based upon the assay described by Cepek, K., et al., in J. Immunol. 150(8):3459-3470 (1993), the entire contents of which are incorporated herein by reference.

The 16E6.A5 cell line was derived by immortalization of the 76N normal epithelial cell line through transfection of the E6 and E7 genes of human papilloma virus (by V. Band, Tufts University, publically available). Monolayers of these adherent cells are grown in flat-bottomed 96-well Linbro tissue culture plates.  $10^4$  adherent cells in 100  $\mu$ l complete media are added per well and allowed to grow for 3 days until they reach confluence. Just before the addition of iIEL, these cells are washed with assay media. To label iIEL, 25 $\mu$ g of 2',7'-bis-(2-carboxyethyl)-5 (and 6)-carboxyfluorescein (BCEFC-AM, Molecular Probes, Inc., Eugene, OR) is diluted in 5  $\mu$ l DMSO and added to a suspension of  $5 \times 10^6$ /ml IEL in complete Yssel's media. The cells are incubated at 37 degrees C for 35 min then washed twice in assay media (phosphate buffered saline "PBS" or RPMI tissue culture media, available from GIBCO or Sigma Chemical Corp., St. Louis, MO, containing 1 mM  $\text{CaCl}_2$ , 2 mM  $\text{MgCl}_2$ , and 10 mM HEPES). After washing, 50,000 labeled iIEL in 100  $\mu$ l of assay media are added to the adherent cell monolayers. iIEL are allowed to settle onto adherent cell monolayers for 25 min or 40 min at 37 degrees C. Unbound cells are removed by flicking media from the plate. Bound cells are detected using a Fluorescence plate reader (IDEXX Co., Portland ME). If antibody blocking is performed, the iIEL are preincubated with an appropriate concentration of antibody (e.g., a 1/250 dilution of ascites fluid or 10  $\mu$ g/ml of purified mAb) for 5 min at 37 degrees C before addition to the adherent cell monolayers. At least four replicates were performed. The % cells bound is calculated by reading the fluorescence units obtained after unbound cells are washed off and dividing this number by the input fluorescence units obtained after adding 50,000 cells/well and multiplying by 100. Serial dilutions of labeled cells have shown that as few as 1000 cells can be detected in the linear range.

To screen a molecular library or other mixture for the



presence of a functionally equivalent peptide analog, the iIEL are washed with HBSS (Hanks buffered saline solution, available from Gibco) and pre-equilibrated with HBSS containing serial dilutions of the library or other peptide-containing solution (over a broad concentration range (e.g. 1 ng/ml to 100 ug/ml) for selected times (e.g., 30 min, 1 hour, 2 hours, 6 hours) at 37 degree C before incubation with 16E6.A5 monolayers that have been washed with HBSS. Functionally equivalent peptide analogs are identified by their ability to inhibit the binding of cells to the monolayer of adherent cells.

Example 3. Construction of a chimeric mouse containing a mutated  $\alpha^F$  gene

In order to produce a mouse in which the  $\alpha^F$  encoding gene had been mutated, a construct was generated to target disruption of the  $\alpha^F$  encoding gene in embryonic stem (ES) cells, and a strategy was devised to identify which of the resulting clones had integrated the construct by homologous recombination (see Figures 6 and 7). The restriction endonuclease sites used to prepare the knockout construct are shown in Figures 6 and 7. The fragments from the genomic clone used in the construction of the knockout construct are indicated by the thicker black lines, and the neomycin gene between genomic clones by the dashed lines. The probes planned for use in Southern blot analysis of genomic are indicated by shaded lines.

The construct was transfected by electroporation into the embryonic stem (ES) cell line D3. The resulting ES cell line was selected with the antibiotic G418, and then cloned by limiting dilution. DNA then was isolated from each clone, subjected to restriction endonuclease digestion, and analyzed by Southern blotting as follows. The restriction endonuclease digestion strategy which is used to identify homologous recombination within genomic DNA isolated from ES cells or mouse tail biopsies is shown in Figure 7. The DNA is digested with either Neo I (left) or Bam HI (right). The length of the generated fragments in the wild type (top) or recombined (bottom)  $\alpha^F$  genes are indicated by arrows above and below the schematic restriction endonuclease fragments. The neomycin resistance gene used to replace the Xho I/Hind III

fragment is shown by a dashed line, and the additional restriction endonuclease site introduced by the neomycin resistance gene are shown below the lower schematics. The probes are shown by shaded thick lines below the schematic.

Two of the clones which integrated the  $\alpha^E$  construct into the genome by homologous recombination were injected into C5/BL/6J blastocysts. These blastocytes then were implanted into C57BL/6J female mice to produce chimeric animals. The chimeric animals were bred, and the F1 animals analyzed by Southern blot analysis of tail genomic DNA to identify mice which transmitted the disrupted  $\alpha^E$  gene to the offspring. The heterozygous animals were mated to yield (129/Sv x C57BL/6J)F2 animals, some of which were homozygous for the disrupted  $\alpha^E$  gene ( $\alpha^E$ /null). In the animals analyzed to date, the homozygous  $\alpha^E$ /null animals had a reduced number of intraepithelial lymphocyte (60-70/1000 epithelial cells versus 150-220/1000 epithelial cells in heterozygous and homozygous wild-type animals). This observation supports the hypothesis that  $\alpha^E$  is important in intraepithelial lymphocyte localization and/or function.

It should be understood that the preceding is merely a detailed description of certain preferred embodiments. It therefore should be apparent to those skilled in the art that various modifications and equivalents can be made without departing from the spirit and scope of the invention.

A Sequence Listing is presented below and is followed by what is claimed.

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT:
  - (A) NAME: Brigham and Women's Hospital, Inc.
  - (B) STREET: 75 Francis Street
  - (C) CITY: Boston
  - (D) STATE: MA
  - (E) COUNTRY: USA
  - (F) ZIP: 02215
- (ii) TITLE OF INVENTION: Novel integrin alpha subunit
- (iii) NUMBER OF SEQUENCES: 25
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Wolf, Greenfield and Sacks, P.C.
  - (B) STREET: 600 Atlantic Avenue
  - (C) CITY: Boston
  - (D) STATE: MA
  - (E) COUNTRY: USA
  - (F) ZIP: 02210
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: not available
  - (B) FILING DATE: herewith
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NO. 08/199,776
  - (B) FILING DATE: 18 February 1994
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Plumer, Elizabeth R.
  - (B) REGISTRATION NUMBER: 36,637
  - (C) REFERENCE/DOCKET NUMBER: B0801/7020WO
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: 617-720-3500
  - (B) TELEFAX: 617-720-2441

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 3933 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA to mRNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) ORIGINAL SOURCE:
  - (A) ORGANISM: Homo sapiens
  - (G) CELL TYPE: mucosal lymphocyte

## (ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 126..3662

## (ix) FEATURE:

(A) NAME/KEY: mat\_peptide

(B) LOCATION: 180..3659

## (ix) FEATURE:

(A) NAME/KEY: signal peptide

(B) LOCATION: 126..179

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GAATTCGGGC CCCCGTGTCT GGGCGTCCGC CTCCTGGCCT CCTGGCTGAG GGGAAAGCTGA	60
GTGGGCCACG GCCCATGTGT CGCACTCGCC TCGGCTCCCA CACAGCCGCC TCTGCTCCAG	120
CAAGG ATG TGG CTC TTC CAC ACT CTG CTC TGC ATA GCC AGC CTG GCC	167
Met Trp Leu Phe His Thr Leu Leu Cys Ile Ala Ser Leu Ala	
-18 -15 -10 -5	
CTG CTG GCC GCT TTC AAT GTG GAT GTG GCC CGG CCC TGG CTC ACG CCC	215
Leu Leu Ala Ala Phe Asn Val Asp Val Ala Arg Pro Trp Leu Thr Pro	
1 5 10	
AAG GGA GGT GCC CCT TTC GTG CTC AGC TCC CTT CTG CAC CAA GAC CCC	263
Lys Gly Gly Ala Pro Phe Val Leu Ser Ser Leu Leu His Gln Asp Pro	
15 20 25	
AGC ACC AAC CAG ACC TGG CTC CTG GTC ACC AGC CCC AGA ACC AAG AGG	311
Ser Thr Asn Gln Thr Trp Leu Leu Val Thr Ser Pro Arg Thr Lys Arg	
30 35 40	
ACA CCA GGG CCC CTC CAT CGA TGT TCC CTT GTC CAG GAT GAA ATC CTT	359
Thr Pro Gly Pro Leu His Arg Cys Ser Leu Val Gln Asp Glu Ile Leu	
45 50 55 60	
TGC CAT CCT GTA GAG CAT GTC CCC ATC CAA GGG GAG GCA CCG GGG AGT	407
Cys His Pro Val Glu His Val Pro Ile Gln Gly Glu Ala Pro Gly Ser	
65 70 75	
GAC CGT TGT CCG GAG CCA CCA CGG TGT TTT GAT ATG CAT TCA AGT GCT	455
Asp Arg Cys Pro Glu Pro Pro Arg Cys Phe Asp Met His Ser Ser Ala	
80 85 90	
GGT CCG GCG CCT CAC AGC CTC AGC TCA GAA CTC ACA GGC ACC TGT AGC	503
Gly Pro Ala Pro His Ser Leu Ser Ser Glu Leu Thr Gly Thr Cys Ser	
95 100 105	
CTC CTG GGC CCT GAC CTC CGT CCC CAG GCT CAG GCC AAC TTC TTC GAC	551
Leu Leu Gly Pro Asp Leu Arg Pro Gln Ala Gln Ala Asn Phe Phe Asp	
110 115 120	
CTT GAA AAT CTC CTG GAT CCA GAT GCA CGT GTG GAC ACT GGA GAC TGC	599
Leu Glu Asn Leu Leu Asp Pro Asp Ala Arg Val Asp Thr Gly Asp Cys	
125 130 135 140	

TAC Tyr	AGC Ser	AAC Asn	AAA Lys	GAA Glu 145	GGC Gly	GGT Gly	GGA Gly	GAA Glu	GAC Asp 150	GAT Asp	GTG Val	AAC Asn	ACA Thr	GCC Ala 155	AGG Arg	647
CAG Gln	CGC Arg	CGG Arg	GCT Ala 160	CTG Leu	GAG Glu	AAG Lys	GAG Glu	GAG Glu	GAG Glu	GAA Glu	GAC Asp	AAG Lys	GAG Glu 170	GAG Glu	GAG Glu	695
GAA Glu	GAC Asp 175	GAG Glu	GAG Glu	GAG Glu	GAG Glu	GAA Glu	GCT Ala 180	GGC Gly	ACC Thr	GAG Glu	ATT Ile	GCC Ala 185	ATC Ile	ATC Ile	CTG Leu	743
GAT Asp 190	GGC Gly	TCA Ser	GGA Gly	AGC Ser	ATT Ile	GAT Asp 195	CCC Pro	CCA Pro	GAC Asp	TTT Phe	CAG Gln 200	AGA Arg	GCC Ala	AAA Lys	GAC Asp	791
TTC Phe 205	ATC Ile	TCC Ser	AAC Asn	ATG Met 210	ATG Met	AGG Arg	AAC Asn	TTC Phe	TAT Tyr	GAA Glu 215	AAG Lys	TGT Cys	TTT Phe	GAG Glu	TGC Cys 220	839
AAC Asn	TTT Phe	GCC Ala	TTG Leu	GTG Val 225	CAG Gln	TAT Tyr	GGA Gly	GGA Gly	GTG Val 230	ATC Ile	CAG Gln	ACT Thr	GAG Glu	TTT Phe 235	GAC Asp	887
CTT Leu	CGG Arg	GAC Asp	AGC Ser 240	CAG Gln	GAT Asp	GTG Val	ATG Met	GCC Ala 245	TCC Ser	CTC Leu	GCC Ala	AGA Arg	GTC Val 250	CAG Gln	AAC Asn	935
ATC Ile	ACT Thr	CAA Gln 255	GTG Val	GGG Gly	AGT Ser	GTC Val 260	ACC Thr	AAG Lys	ACT Thr	GCC Ala	TCA Ser	GCC Ala 265	ATG Met	CAA Gln	CAC His	983
GTC Val 270	TTA Leu	GAC Asp	AGC Ser	ATC Ile	TTC Phe	ACC Thr 275	TCA Ser	AGC Ser	CAC His	GGC Gly	TCC Ser	AGG Arg	AGA Arg	AAG Lys	GCA Ala	1031
TCC Ser 285	AAG Lys	GTC Val	ATG Met	GTG Val	GTG Val	CTC Leu 290	ACC Thr	GAT Asp	GGT Gly	GGC Gly 295	ATA Ile	TTC Phe	GAG Glu	GAC Asp	CCC Pro 300	1079
CTC Leu	AAC Asn	CTT Leu	ACG Thr	ACA Thr 305	GTC Val	ATC Ile	AAC Asn	TCC Ser	CCC Pro 310	AAA Lys	ATG Met	CAG Gln	GGT Gly 315	GTT Val	GAG Glu	1127
CGC Arg	TTT Phe	GCC Ala	ATT Ile 320	GGG Gly	GTG Val	GGA Gly	GAA Glu	GAA Glu	TTT Phe 325	AAG Lys	AGT Ser	GCT Ala	AGG Arg 330	ACT Thr	GCG Ala	1175
AGG Arg	GAA Glu	CTG Leu	AAC Asn 335	CTG Leu	ATC Ile	GCC Ala	TCA Ser 340	GAC Asp	CCG Pro	GAT Asp	GAG Glu	ACC Thr 345	CAT His	GCT Ala	TTC Phe	1223
AAG Lys 350	GTG Val	ACC Thr	AAC Asn	TAC Tyr	ATG Met	GCG Ala 355	CTG Leu	GAT Asp	GGG Gly	CTG Leu	CTG Leu	AGC Ser	AAA Lys	CTG Leu	CGG Arg	1271

TAC Tyr 365	AAC Asn	ATC Ile	ATC Ile	AGC Ser	ATG Met	GAA Glu	GGC Gly	ACG Thr	GTT Val	GGG Gly	GAC Asp	GCC Ala	CTT Leu	CAC His	TAC Tyr	1319
CAG Gln	CTG Leu	GCA Ala	CAG Gln	ATT Ile	GGC Gly	TTC Phe	AGT Ser	GCT Ala	CAG Gln	ATC Ile	CTG Leu	GAT Asp	GAG Glu	CGG Arg	CAG Gln	1367
GTG Val	CTG Leu	CTC Leu	GGC Gly	GCC Ala	GTC Val	GGG Gly	GCC Ala	TTT Phe	GAC Asp	TGG Trp	TCC Ser	GGA Gly	GGG Gly	GCG Ala	TTG Leu	1415
CTC Leu	TAC Tyr	GAC Asp	ACA Thr	CGC Arg	AGC Ser	CGC Arg	CGG Arg	GGC Gly	CGC Arg	TTC Phe	CTG Leu	AAC Asn	CAG Gln	ACA Thr	GCG Ala	1463
GCG Ala	GCG Ala	GCG Ala	GCA Ala	GAC Asp	GCG Ala	GAG Glu	GCT Ala	GCG Ala	CAG Gln	TAC Tyr	AGC Ser	TAC Tyr	CTG Leu	GGT Gly	TAC Tyr	1511
GCT Ala	GTG Val	GCC Ala	GTG Val	CTG Leu	CAC His	AAG Lys	ACC Thr	TGC Cys	AGC Ser	CTC Leu	TCC Ser	TAC Tyr	GTC Val	GCG Ala	GGG Gly	1559
GCT Ala	CCA Pro	CAG Gln	TAC Tyr	AAA Lys	CAT His	CAT His	GGG Gly	GCC Ala	GTG Val	TTT Phe	GAG Glu	CTC Leu	CAG Gln	AAG Lys	GAG Glu	1607
GGC Gly	AGA Arg	GAG Glu	GCC Ala	AGC Ser	TTC Phe	CTG Leu	CCA Pro	GTG Val	CTG Leu	GAG Glu	GGA Gly	GAG Glu	CAG Gln	ATG Met	GGG Gly	1655
TCC Ser	TAT Tyr	TTT Phe	GGC Gly	TCT Ser	GAG Glu	CTG Leu	TGC Cys	CCT Pro	GTG Val	GAC Asp	ATT Ile	GAC Asp	ATG Met	GAT Asp	GGA Gly	1703
AGC Ser	ACG Thr	GAC Asp	TTC Phe	TTG Leu	CTG Leu	GTG Val	GCT Ala	GCT Ala	CCA Pro	TTT Phe	TAC Tyr	CAC His	GTT Val	CAT His	GGA Gly	1751
GAA Glu	GAA Glu	GGC Gly	AGA Arg	GTC Val	TAC Tyr	GTG Val	TAC Tyr	CGT Arg	CTC Leu	AGC Ser	GAG Glu	CAG Gln	GAT Asp	GGT Gly	TCT Ser	1799
TTC Phe	TCC Ser	TTG Leu	GCA Ala	CGC Arg	ATA Ile	CTG Leu	AGT Ser	GGG Gly	CAC His	CCC Pro	GGG Gly	TTC Phe	ACC Thr	AAT Asn	GCC Ala	1847
CGC Arg	TTT Phe	GGC Gly	TTT Phe	GCC Ala	ATG Met	GCG Ala	GCT Ala	ATG Met	GGG Gly	GAT Asp	CTC Leu	AGT Ser	CAG Gln	GAT Asp	AAG Lys	1895
CTC Leu	ACA Thr	GAT Asp	GTG Val	GCC Ala	ATC Ile	GGG Gly	GCC Ala	CCC Pro	CTG Leu	GAA Glu	GGT Gly	TTT Phe	GGG Gly	GCA Ala	GAT Asp	1943

GAT Asp	GGT Gly	GCC Ala	AGC Ser	TTC Phe	GGC Gly	AGT Ser	GTG Val	TAT Tyr	ATC Ile	TAC Tyr	AAT Asn	GGA Gly	CAC His	TGG Trp	GAC Asp	1991
590						595					600					
GGC Gly	CTC Leu	TCC Ser	GCC Ala	AGC Ser	CCC Pro	TCG Ser	CAG Gln	CGG Arg	ATC Ile	AGA Arg	GCC Ala	TCC Ser	ACG Thr	GTG Val	GCC Ala	2039
605					610					615					620	
CCA Pro	GGA Gly	CTC Leu	CAG Gln	TAC Tyr	TTC Phe	GGC Gly	ATG Met	TCC Ser	ATG Met	GCT Ala	GGT Gly	GGC Gly	TTT Phe	GAT Asp	ATT Ile	2087
				625					630					635		
AGT Ser	GGC Gly	GAC Asp	GGC Gly	CTT Leu	GCC Ala	GAC Asp	ATC Ile	ACC Thr	GTG Val	GGC Gly	ACT Thr	CTG Leu	GGC Gly	CAG Gln	GCG Ala	2135
			640					645					650			
GTT Val	GTG Val	TTC Phe	CGC Arg	TCC Ser	CGG Arg	CCT Pro	GTG Val	GTT Val	CGC Arg	CTG Leu	AAG Lys	GTC Val	TCC Ser	ATG Met	GCC Ala	2183
		655					660					665				
TTC Phe	ACC Thr	CCC Pro	AGC Ser	GCA Ala	CTG Leu	CCC Pro	ATC Ile	GGC Gly	TTC Phe	AAC Asn	GGC Gly	GTC Val	GTG Val	AAT Asn	GTC Val	2231
	670					675					680					
CGT Arg	TTA Leu	TGT Cys	TTT Phe	GAA Glu	ATC Ile	AGC Ser	TCT Ser	GTA Val	ACC Thr	ACA Thr	GCC Ala	TCT Ser	GAG Glu	TCA Ser	GGC Gly	2279
685					690					695					700	
CTC Leu	CGT Arg	GAG Glu	GCA Ala	CTT Leu	CTC Leu	AAC Asn	TTC Phe	ACG Thr	CTG Leu	GAT Asp	GTG Val	GAT Asp	GTG Val	GGG Gly	AAG Lys	2327
				705					710					715		
CAG Gln	AGG Arg	AGA Arg	CGG Arg	CTG Leu	CAG Gln	TGT Cys	TCA Ser	GAC Asp	GTA Val	AGA Arg	AGC Ser	TGT Cys	CTG Leu	GGC Gly	TGC Cys	2375
			720					725					730			
CTG Leu	AGG Arg	GAG Glu	TGG Trp	AGC Ser	AGC Ser	GGA Gly	TCC Ser	CAG Gln	CTT Leu	TGT Cys	GAG Glu	GAC Asp	CTC Leu	CTG Leu	CTC Leu	2423
		735					740					745				
ATG Met	CCC Pro	ACA Thr	GAG Glu	GGA Gly	GAG Glu	CTC Leu	TGT Cys	GAG Glu	GAG Glu	GAC Asp	TGC Cys	TTC Phe	TCC Ser	AAT Asn	GCC Ala	2471
		750				755					760					
AGT Ser	GTC Val	AAA Lys	GTC Val	AGC Ser	TAC Tyr	CAG Gln	CTC Leu	CAG Gln	ACC Thr	CCT Pro	GAG Glu	GGA Gly	CAG Gln	ACG Thr	GAC Asp	2519
	765				770					775					780	
CAT His	CCC Pro	CAG Gln	CCC Pro	ATC Ile	CTG Leu	GAC Asp	CGC Arg	TAC Tyr	ACT Thr	GAG Glu	CCC Pro	TTT Phe	GCC Ala	ATC Ile	TTC Phe	2567
				785					790					795		
CAG Gln	CTG Leu	CCC Pro	TAT Tyr	GAG Glu	AAG Lys	GCC Ala	TGC Cys	AAG Lys	AAT Asn	AAG Lys	CTG Leu	TTT Phe	TGT Cys	GTC Val	GCA Ala	2615
			800					805					810			

GAA TTA CAG TTG GCC ACC ACC GTC TCT CAG CAG GAG TTG GTG GTG GGT Glu Leu Gln Leu Ala Thr Thr Val Ser Gln Gln Glu Leu Val Val Gly 815 820 825	2663
CTC ACA AAG GAG CTG ACC CTG AAC ATT AAC CTA ACT AAC TCC GGG GAA Leu Thr Lys Glu Leu Thr Leu Asn Ile Asn Leu Thr Asn Ser Gly Glu 830 835 840	2711
GAT TCC TAC ATG ACA AGC ATG GCC TTG AAT TAC CCC AGA AAC CTG CAG Asp Ser Tyr Met Thr Ser Met Ala Leu Asn Tyr Pro Arg Asn Leu Gln 845 850 855 860	2759
TTG AAG AGG ATG CAA AAG CCT CCC TCT CCA AAC ATT CAG TGT GAT GAC Leu Lys Arg Met Gln Lys Pro Pro Ser Pro Asn Ile Gln Cys Asp Asp 865 870 875	2807
CCT CAG CCG GTT GCT TCT GTC CTG ATC ATG AAC TGC AGG ATT GGT CAC Pro Gln Pro Val Ala Ser Val Leu Ile Met Asn Cys Arg Ile Gly His 880 885 890	2855
CCC GTC CTC AAG AGG TCA TCT GCT CAT GTT TCA GTC GTT TGG CAG CTA Pro Val Leu Lys Arg Ser Ser Ala His Val Ser Val Val Trp Gln Leu 895 900 905	2903
GAG GAG AAT GCC TTT CCA AAC AGG ACA GCA GAC ATC ACT GTG ACT GTC Glu Glu Asn Ala Phe Pro Asn Arg Thr Ala Asp Ile Thr Val Thr Val 910 915 920	2951
ACC AAT TCC AAT GAA AGA CGG TCT TTG GCC AAC GAG ACC CAC ACC CTT Thr Asn Ser Asn Glu Arg Arg Ser Leu Ala Asn Glu Thr His Thr Leu 925 930 935 940	2999
CAA TTC AGG CAT GGC TTC GTT GCA GTT CTG TCC AAA CCA TCC ATA ATG Gln Phe Arg His Gly Phe Val Ala Val Leu Ser Lys Pro Ser Ile Met 945 950 955	3047
TAC GTG AAC ACA GGC CAG GGG CTT TCT CAC CAC AAA GAA TTC CTC TTC Tyr Val Asn Thr Gly Gln Gly Leu Ser His His Lys Glu Phe Leu Phe 960 965 970	3095
CAT GTA CAT GGG GAG AAC CTC TTT GGA GCA GAA TAC CAG TTG CAA ATT His Val His Gly Glu Asn Leu Phe Gly Ala Glu Tyr Gln Leu Gln Ile 975 980 985	3143
TGC GTC CCA ACC AAA TTA CGA GGT CTC CAG GTT GCA GCA GTG AAG AAG Cys Val Pro Thr Lys Leu Arg Gly Leu Gln Val Ala Ala Val Lys Lys 990 995 1000	3191
CTG ACG AGG ACT CAG GCC TCC ACG GTG TGC ACC TGG AGT CAG GAG CGC Leu Thr Arg Thr Gln Ala Ser Thr Val Cys Thr Trp Ser Gln Glu Arg 1005 1010 1015 1020	3239
GCT TGT GCG TAC AGT TCG GTT CAG CAT GTG GAA GAA TGG CAT TCA GTG Ala Cys Ala Tyr Ser Ser Val Gln His Val Glu Glu Trp His Ser Val 1025 1030 1035	3287



AGC TGT GTC ATC GCT TCA GAT AAA GAA AAT GTC ACC GTG GCT GCA GAG Ser Cys Val Ile Ala Ser Asp Lys Glu Asn Val Thr Val Ala Ala Glu 1040 1045 1050	3335
ATC TCC TGG GAT CAC TCT GAG GAG TTA CTA AAA GAT GTA ACT GAA CTG Ile Ser Trp Asp His Ser Glu Glu Leu Leu Lys Asp Val Thr Glu Leu 1055 1060 1065	3383
CAG ATC CTT GGT GAA ATA TCT TTC AAC AAA TCT CTA TAT GAG GGA CTG Gln Ile Leu Gly Glu Ile Ser Phe Asn Lys Ser Leu Tyr Glu Gly Leu 1070 1075 1080	3431
AAT GCA GAG AAC CAC AGA ACT AAG ATC ACT GTC GTC TTC CTG AAA GAT Asn Ala Glu Asn His Arg Thr Lys Ile Thr Val Val Phe Leu Lys Asp 1085 1090 1095 1100	3479
GAG AAG TAC CAT TCT TTG CCT ATC ATC ATT AAA GGC AGC GTT GGT GGA Glu Lys Tyr His Ser Leu Pro Ile Ile Ile Lys Gly Ser Val Gly Gly 1105 1110 1115	3527
CTT CTG GTG TTG ATC GTG ATT CTG GTC ATC CTG TTC AAG TGT GGC TTT Leu Leu Val Leu Ile Val Ile Leu Val Ile Leu Phe Lys Cys Gly Phe 1120 1125 1130	3575
TTT AAA AGA AAA TAT CAA CAA CTG AAC TTG GAG AGC ATC AGG AAG GCC Phe Lys Arg Lys Tyr Gln Gln Leu Asn Leu Glu Ser Ile Arg Lys Ala 1135 1140 1145	3623
CAG CTG AAA TCA GAG AAT CTG CTC GAA GAA GAG AAT TAGGACCTGC Gln Leu Lys Ser Glu Asn Leu Leu Glu Glu Glu Asn 1150 1155 1160	3669
TATCCACTGG GAGAGGCTAT CAGCCAGTCC TGGGACTTGG AGACCCAGCA TCCTTTGCAT	3729
TACTTTTTCC TTCAGGATGA TCTAGAGCAG CATGGAGCTG TTGGTAGAAT ATTAGTTTTT	3789
AACCATACAT TGTCCTGTGC ATTGTGCAAA AAGTAACTT AGGAAACATT	3849
TGGTATTAATA TAAATTTACA CTTTCTTTG CAAAAAAAAA AAAAAAAAAA AAAAAAAAAA	3909
AAAAAAAAAA AAAAACCAGA ATTC	3933

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1178 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Trp Leu Phe His Thr Leu Leu Cys Ile Ala Ser Leu Ala Leu Leu  
-18 -15 -10 -5

Ala Ala Phe Asn Val Asp Val Ala Arg Pro Trp Leu Thr Pro Lys Gly  
                   1                                  5                                  10  
 Gly Ala Pro Phe Val Leu Ser Ser Leu Leu His Gln Asp Pro Ser Thr  
   15                                  20                                  25                                  30  
 Asn Gln Thr Trp Leu Leu Val Thr Ser Pro Arg Thr Lys Arg Thr Pro  
                                   35                                  40                                  45  
 Gly Pro Leu His Arg Cys Ser Leu Val Gln Asp Glu Ile Leu Cys His  
                                   50                                  55                                  60  
 Pro Val Glu His Val Pro Ile Gln Gly Glu Ala Pro Gly Ser Asp Arg  
                                   65                                  70                                  75  
 Cys Pro Glu Pro Pro Arg Cys Phe Asp Met His Ser Ser Ala Gly Pro  
   80                                  85                                  90  
 Ala Pro His Ser Leu Ser Ser Glu Leu Thr Gly Thr Cys Ser Leu Leu  
   95                                  100                                  105                                  110  
 Gly Pro Asp Leu Arg Pro Gln Ala Gln Ala Asn Phe Phe Asp Leu Glu  
                                   115                                  120                                  125  
 Asn Leu Leu Asp Pro Asp Ala Arg Val Asp Thr Gly Asp Cys Tyr Ser  
                                   130                                  135                                  140  
 Asn Lys Glu Gly Gly Gly Glu Asp Asp Val Asn Thr Ala Arg Gln Arg  
                                   145                                  150                                  155  
 Arg Ala Leu Glu Lys Glu Glu Glu Glu Asp Lys Glu Glu Glu Glu Asp  
   160                                  165                                  170  
 Glu Glu Glu Glu Glu Ala Gly Thr Glu Ile Ala Ile Ile Leu Asp Gly  
   175                                  180                                  185                                  190  
 Ser Gly Ser Ile Asp Pro Pro Asp Phe Gln Arg Ala Lys Asp Phe Ile  
                                   195                                  200                                  205  
 Ser Asn Met Met Arg Asn Phe Tyr Glu Lys Cys Phe Glu Cys Asn Phe  
                                   210                                  215                                  220  
 Ala Leu Val Gln Tyr Gly Gly Val Ile Gln Thr Glu Phe Asp Leu Arg  
                                   225                                  230                                  235  
 Asp Ser Gln Asp Val Met Ala Ser Leu Ala Arg Val Gln Asn Ile Thr  
   240                                  245                                  250  
 Gln Val Gly Ser Val Thr Lys Thr Ala Ser Ala Met Gln His Val Leu  
   255                                  260                                  265                                  270  
 Asp Ser Ile Phe Thr Ser Ser His Gly Ser Arg Arg Lys Ala Ser Lys  
                                   275                                  280                                  285  
 Val Met Val Val Leu Thr Asp Gly Gly Ile Phe Glu Asp Pro Leu Asn  
                                   290                                  295                                  300

Leu Thr Thr Val Ile Asn Ser Pro Lys Met Gln Gly Val Glu Arg Phe  
 305 310 315  
 Ala Ile Gly Val Gly Glu Glu Phe Lys Ser Ala Arg Thr Ala Arg Glu  
 320 325 330  
 Leu Asn Leu Ile Ala Ser Asp Pro Asp Glu Thr His Ala Phe Lys Val  
 335 340 345 350  
 Thr Asn Tyr Met Ala Leu Asp Gly Leu Leu Ser Lys Leu Arg Tyr Asn  
 355 360 365  
 Ile Ile Ser Met Glu Gly Thr Val Gly Asp Ala Leu His Tyr Gln Leu  
 370 375 380  
 Ala Gln Ile Gly Phe Ser Ala Gln Ile Leu Asp Glu Arg Gln Val Leu  
 385 390 395  
 Leu Gly Ala Val Gly Ala Phe Asp Trp Ser Gly Gly Ala Leu Leu Tyr  
 400 405 410  
 Asp Thr Arg Ser Arg Arg Gly Arg Phe Leu Asn Gln Thr Ala Ala Ala  
 415 420 425 430  
 Ala Ala Asp Ala Glu Ala Ala Gln Tyr Ser Tyr Leu Gly Tyr Ala Val  
 435 440 445  
 Ala Val Leu His Lys Thr Cys Ser Leu Ser Tyr Val Ala Gly Ala Pro  
 450 455 460  
 Gln Tyr Lys His His Gly Ala Val Phe Glu Leu Gln Lys Glu Gly Arg  
 465 470 475  
 Glu Ala Ser Phe Leu Pro Val Leu Glu Gly Glu Gln Met Gly Ser Tyr  
 480 485 490  
 Phe Gly Ser Glu Leu Cys Pro Val Asp Ile Asp Met Asp Gly Ser Thr  
 495 500 505 510  
 Asp Phe Leu Leu Val Ala Ala Pro Phe Tyr His Val His Gly Glu Glu  
 515 520 525  
 Gly Arg Val Tyr Val Tyr Arg Leu Ser Glu Gln Asp Gly Ser Phe Ser  
 530 535 540  
 Leu Ala Arg Ile Leu Ser Gly His Pro Gly Phe Thr Asn Ala Arg Phe  
 545 550 555  
 Gly Phe Ala Met Ala Ala Met Gly Asp Leu Ser Gln Asp Lys Leu Thr  
 560 565 570  
 Asp Val Ala Ile Gly Ala Pro Leu Glu Gly Phe Gly Ala Asp Asp Gly  
 575 580 585 590  
 Ala Ser Phe Gly Ser Val Tyr Ile Tyr Asn Gly His Trp Asp Gly Leu  
 595 600 605

Ser Ala Ser Pro Ser Gln Arg Ile Arg Ala Ser Thr Val Ala Pro Gly  
 610 615 620  
 Leu Gln Tyr Phe Gly Met Ser Met Ala Gly Gly Phe Asp Ile Ser Gly  
 625 630 635  
 Asp Gly Leu Ala Asp Ile Thr Val Gly Thr Leu Gly Gln Ala Val Val  
 640 645 650  
 Phe Arg Ser Arg Pro Val Val Arg Leu Lys Val Ser Met Ala Phe Thr  
 655 660 665 670  
 Pro Ser Ala Leu Pro Ile Gly Phe Asn Gly Val Val Asn Val Arg Leu  
 675 680 685  
 Cys Phe Glu Ile Ser Ser Val Thr Thr Ala Ser Glu Ser Gly Leu Arg  
 690 695 700  
 Glu Ala Leu Leu Asn Phe Thr Leu Asp Val Asp Val Gly Lys Gln Arg  
 705 710 715  
 Arg Arg Leu Gln Cys Ser Asp Val Arg Ser Cys Leu Gly Cys Leu Arg  
 720 725 730  
 Glu Trp Ser Ser Gly Ser Gln Leu Cys Glu Asp Leu Leu Leu Met Pro  
 735 740 745 750  
 Thr Glu Gly Glu Leu Cys Glu Glu Asp Cys Phe Ser Asn Ala Ser Val  
 755 760 765  
 Lys Val Ser Tyr Gln Leu Gln Thr Pro Glu Gly Gln Thr Asp His Pro  
 770 775 780  
 Gln Pro Ile Leu Asp Arg Tyr Thr Glu Pro Phe Ala Ile Phe Gln Leu  
 785 790 795  
 Pro Tyr Glu Lys Ala Cys Lys Asn Lys Leu Phe Cys Val Ala Glu Leu  
 800 805 810  
 Gln Leu Ala Thr Thr Val Ser Gln Gln Glu Leu Val Val Gly Leu Thr  
 815 820 825 830  
 Lys Glu Leu Thr Leu Asn Ile Asn Leu Thr Asn Ser Gly Glu Asp Ser  
 835 840 845  
 Tyr Met Thr Ser Met Ala Leu Asn Tyr Pro Arg Asn Leu Gln Leu Lys  
 850 855 860  
 Arg Met Gln Lys Pro Pro Ser Pro Asn Ile Gln Cys Asp Asp Pro Gln  
 865 870 875  
 Pro Val Ala Ser Val Leu Ile Met Asn Cys Arg Ile Gly His Pro Val  
 880 885 890  
 Leu Lys Arg Ser Ser Ala His Val Ser Val Val Trp Gln Leu Glu Glu  
 895 900 905 910

Asn Ala Phe Pro Asn Arg Thr Ala Asp Ile Thr Val Thr Val Thr Asn  
                             915                            920                            925  
  
 Ser Asn Glu Arg Arg Ser Leu Ala Asn Glu Thr His Thr Leu Gln Phe  
                             930                            935                            940  
  
 Arg His Gly Phe Val Ala Val Leu Ser Lys Pro Ser Ile Met Tyr Val  
                             945                            950                            955  
  
 Asn Thr Gly Gln Gly Leu Ser His His Lys Glu Phe Leu Phe His Val  
                             960                            965                            970  
  
 His Gly Glu Asn Leu Phe Gly Ala Glu Tyr Gln Leu Gln Ile Cys Val  
                             975                            980                            985                            990  
  
 Pro Thr Lys Leu Arg Gly Leu Gln Val Ala Ala Val Lys Lys Leu Thr  
                             995                            1000                            1005  
  
 Arg Thr Gln Ala Ser Thr Val Cys Thr Trp Ser Gln Glu Arg Ala Cys  
                             1010                            1015                            1020  
  
 Ala Tyr Ser Ser Val Gln His Val Glu Glu Trp His Ser Val Ser Cys  
                             1025                            1030                            1035  
  
 Val Ile Ala Ser Asp Lys Glu Asn Val Thr Val Ala Ala Glu Ile Ser  
                             1040                            1045                            1050  
  
 Trp Asp His Ser Glu Glu Leu Leu Lys Asp Val Thr Glu Leu Gln Ile  
                             1055                            1060                            1065                            1070  
  
 Leu Gly Glu Ile Ser Phe Asn Lys Ser Leu Tyr Glu Gly Leu Asn Ala  
                             1075                            1080                            1085  
  
 Glu Asn His Arg Thr Lys Ile Thr Val Val Phe Leu Lys Asp Glu Lys  
                             1090                            1095                            1100  
  
 Tyr His Ser Leu Pro Ile Ile Ile Lys Gly Ser Val Gly Gly Leu Leu  
                             1105                            1110                            1115  
  
 Val Leu Ile Val Ile Leu Val Ile Leu Phe Lys Cys Gly Phe Phe Lys  
                             1120                            1125                            1130  
  
 Arg Lys Tyr Gln Gln Leu Asn Leu Glu Ser Ile Arg Lys Ala Gln Leu  
                             1135                            1140                            1145                            1150  
  
 Lys Ser Glu Asn Leu Leu Glu Glu Glu Asn  
                             1155                            1160

## (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 165 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: cDNA to mRNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: Homo sapien
  - (G) CELL TYPE: mucosal lymphocytes
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```

GAAAATCTCC TGGATCCAGA TGCACGTGTG GACACTGGAG ACTGCTACAG CAACAAAGAA      60
GGCGGTGGAG AAGACGATGT GAACACAGCC AGGCAGCGCC GGGCTCTGGA GAAGGAGGAG      120
GAGGAAGACA AGGAGGAGGA GGAAGACGAG GAGGAGGAGG AAGCT                      165

```

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 55 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:
 

- (A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```

Glu Asn Leu Leu Asp Pro Asp Ala Arg Val Asp Thr Gly Asp Cys Tyr
1           5           10           15
Ser Asn Lys Glu Gly Gly Gly Glu Asp Asp Val Asn Thr Ala Arg Gln
20           25           30
Arg Arg Ala Leu Glu Lys Glu Glu Glu Asp Lys Glu Glu Glu Glu
35           40           45
Asp Glu Glu Glu Glu Glu Ala
50           55

```

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 amino acids

(B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Ala Leu Glu Lys Glu Glu Glu Glu Asp Lys Glu Glu Glu Glu Asp Glu  
 1                      5                      10                      15  
 Glu Glu Glu Glu Ala  
                     20

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 34 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapien

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Glu Asn Leu Leu Asp Pro Asp Ala Arg Val Asp Thr Gly Asp Cys Tyr  
 1                      5                      10                      15  
 Ser Asn Lys Glu Gly Gly Gly Glu Asp Asp Val Asn Thr Ala Arg Gln  
                     20                      25                      30  
 Arg Arg

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 102 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GAAAATCTCC TGGATCCAGA TGCACGTGTG GACACTGGAG ACTGCTACAG CAACAAAGAA	60
GGCGGTGGAG AAGACGATGT GAACACAGCC AGGCAGCGCC GG	102

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Glu Lys Glu Glu Glu Glu Asp Lys Glu Glu
1                    5                    10

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: YES



- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: Homo sapiens
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:  
Lys Glu Glu Glu Glu Asp Lys Glu Glu Glu  
1                    5                    10

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 10 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (iii) HYPOTHETICAL: YES
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: Homo sapiens
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:  
Glu Glu Glu Glu Asp Lys Glu Glu Glu Glu  
1                    5                    10

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 10 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (iii) HYPOTHETICAL: YES
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: Homo sapiens
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Glu Glu Glu Asp Lys Glu Glu Glu Asp  
1 5 10

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 10 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Glu Glu Asp Lys Glu Glu Glu Asp Glu  
1 5 10

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 10 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Glu Asp Lys Glu Glu Glu Asp Glu Glu  
1 5 10

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 10 amino acids

- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Asp Lys Glu Glu Glu Glu Asp Glu Glu Glu  
1 5 10

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Lys Glu Glu Glu Glu Asp Glu Glu Glu Glu  
1 5 10

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: YES

- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: Homo sapiens
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:  
Glu Glu Glu Glu Asp Glu Glu Glu Glu Glu  
1                      5                      10

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 10 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (iii) HYPOTHETICAL: YES
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: Homo sapiens
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:  
Ala Leu Glu Lys Glu Glu Glu Glu Asp Lys  
1                      5                      10

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 14 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (iii) HYPOTHETICAL: YES
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: Homo sapiens
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Glu Glu Asp Lys Glu Glu Glu Glu Asp Glu Glu Glu Glu  
1 5 10

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 14 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Ala Leu Glu Lys Glu Glu Glu Glu Asp Lys Glu Glu Glu  
1 5 10

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 15 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Glu Lys Glu Glu Glu Glu Asp Lys Glu Glu Glu Asp Glu Glu  
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 amino acids

- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Ala Leu Glu Lys Glu Glu Glu Glu Asp Lys Glu Glu Glu Glu Asp Glu  
1                      5                      10                      15  
Glu Glu Glu Glu  
                    20

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:

(A) ORGANISM: synthetic peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Glu Lys Glu Glu Glu Glu Asp Glu Glu Glu Glu  
1                      5                      10

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: synthetic peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Asp Lys Asp Asp Asp Asp Asp Asp Asp Asp Asp Asp  
1 5 10

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 12 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: synthetic peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Glu Asp Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu  
1 5 10

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 13 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: synthetic peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Glu Asn Glu Glu Glu Glu Asp Asn Glu Glu Glu Glu Asp  
1 5 10

CLAIMS

1. An isolated peptide selected from the group consisting of Sequence I.D. Nos. 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 and 25.

2. The peptide of claim 1, wherein the peptide is selected from the group consisting of Sequence I.D. Nos. 4, 5 and 6.

3. A method for screening a molecular library to identify lead compounds which mimic the in vivo activity of an integrin  $\alpha^x$  chain, the method comprising:

determining whether the molecular library contains a compound which inhibits adhesion between a human mucosal lymphocyte-1 antigen and an epithelial cell in vitro.

4. A method for screening a molecular library to identify lead compounds which mimic a ligand-binding site of an integrin  $\alpha^x$  chain, the method comprising:

determining whether the molecular library contains a compound which competitively inhibits binding of a peptide selected from the group consisting of Sequence I.D. Nos. 2, 4, 5 and 6, to an antibody which specifically recognizes the ligand binding site of the  $\alpha^x$  chain.

5. A pharmaceutical composition comprising

a therapeutically effective amount of at least one peptide selected from the group consisting of Sequence I.D. No. 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 and 25; and

a pharmaceutically acceptable carrier therefor.

6. The composition of claim 5, wherein the peptide is selected from the group consisting of Sequence I.D. Nos. 4, 5 and 6.

7. A support having a biologically active surface which exhibits cell attachment activity, said surface having attached thereto at least one peptide selected from the group consisting of Sequence I.D. No. 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 and 25.



8. The support of claim 7, wherein the peptide is selected from the group consisting of Sequence I.D. Nos. 4, 5 and 6.

9. The support of claim 7, wherein the support is selected from the group consisting of a prosthetic device and an affinity matrix.

10. A method for isolating a ligand of the integrin  $\alpha^E$  chain from detergent extracts of cells or cell membranes, the method comprising:

(1) specifically adsorbing the ligand to an affinity matrix having a peptide selected from the group consisting of Sequence I.D. Nos. 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 and 25 coupled thereto to form a ligand-adsorbed affinity matrix; and

(2) adding to the ligand-adsorbed affinity matrix a plurality of peptides selected from the group consisting of Sequence I.D. Nos. 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 and 25 to specifically elute the ligand from the affinity matrix.

11. An antibody specific for an antigen selected from the group consisting of Sequence I.D. Nos. 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 and 25.

12. An isolated oligonucleotide encoding a peptide selected from the group consisting of Sequence I.D. Nos. 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 and 25.

13. An isolated oligonucleotide capable of hybridizing under stringent conditions to the nucleotide sequence residing between positions 555 and 656 inclusive of Sequence I.D. No. 1.

14. The isolated oligonucleotide of claim 13, wherein the oligonucleotide has 100% sequence homology with Sequence I.D. No. 7.

15. An antisense oligonucleotide capable of hybridizing under stringent conditions to the isolated oligonucleotide of claim 13.

16. An expression vector comprising at least one strand of the oligonucleotide of claim 13.

17. The expression vector of claim 16, wherein the at least one strand is operatively joined to a regulatory sequence.

18. A cell line transfected with the recombinant expression vector of claim 16.

19. A pharmaceutical composition comprising:  
a therapeutically effective amount of an isolated oligonucleotide capable of hybridizing under stringent conditions to Sequence I.D. No. 5; and  
a pharmaceutically acceptable carrier therefor.

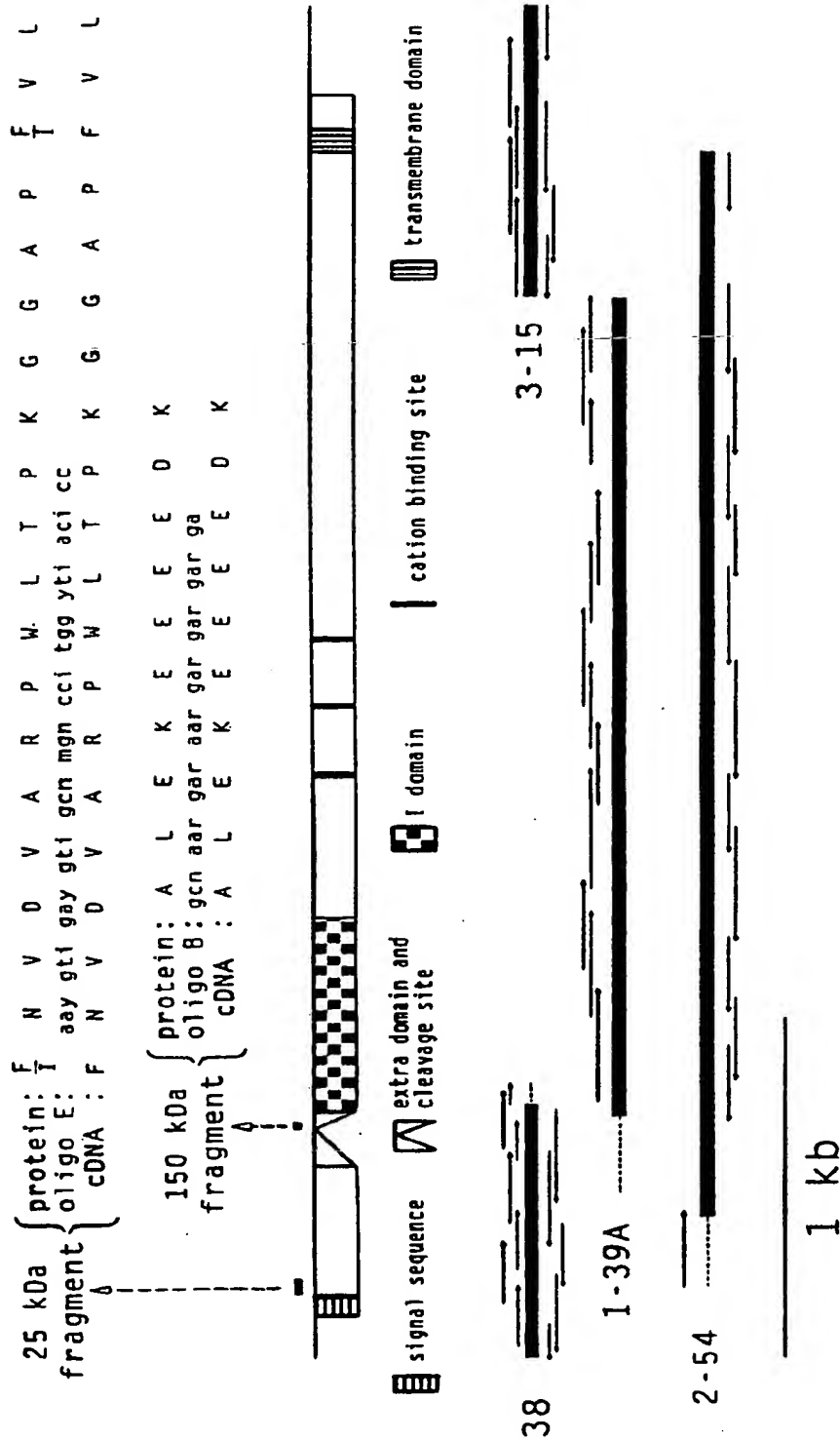


FIG. 1

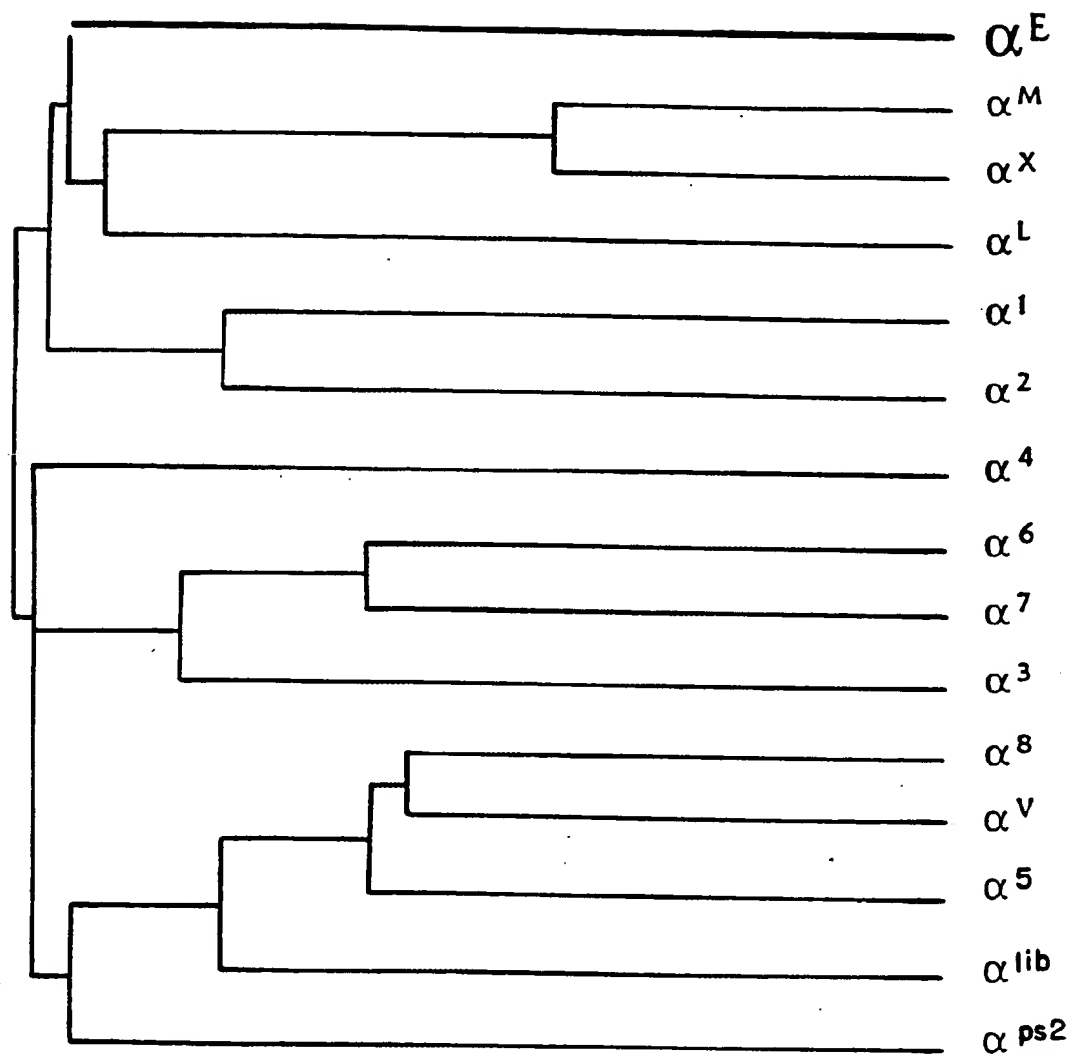


FIG. 2

PERCENT SIMILARITY TO  $\alpha^E$ 

I/X	$\alpha^L$	$\alpha^M$	$\alpha^X$	$\alpha^1$	$\alpha^2$	$\alpha^3$	$\alpha^4$	$\alpha^5$	$\alpha^6$	$\alpha^7$	$\alpha^8$	$\alpha^V$	$\alpha^{lib}$	$\alpha^{ps2}$
removed	52	51	52	51	51	50	50	49	51	49	47	48	47	46
full length	53	52	54	51	52	49	50	49	52	48	46	49	48	42

FIG. 3  
2/7

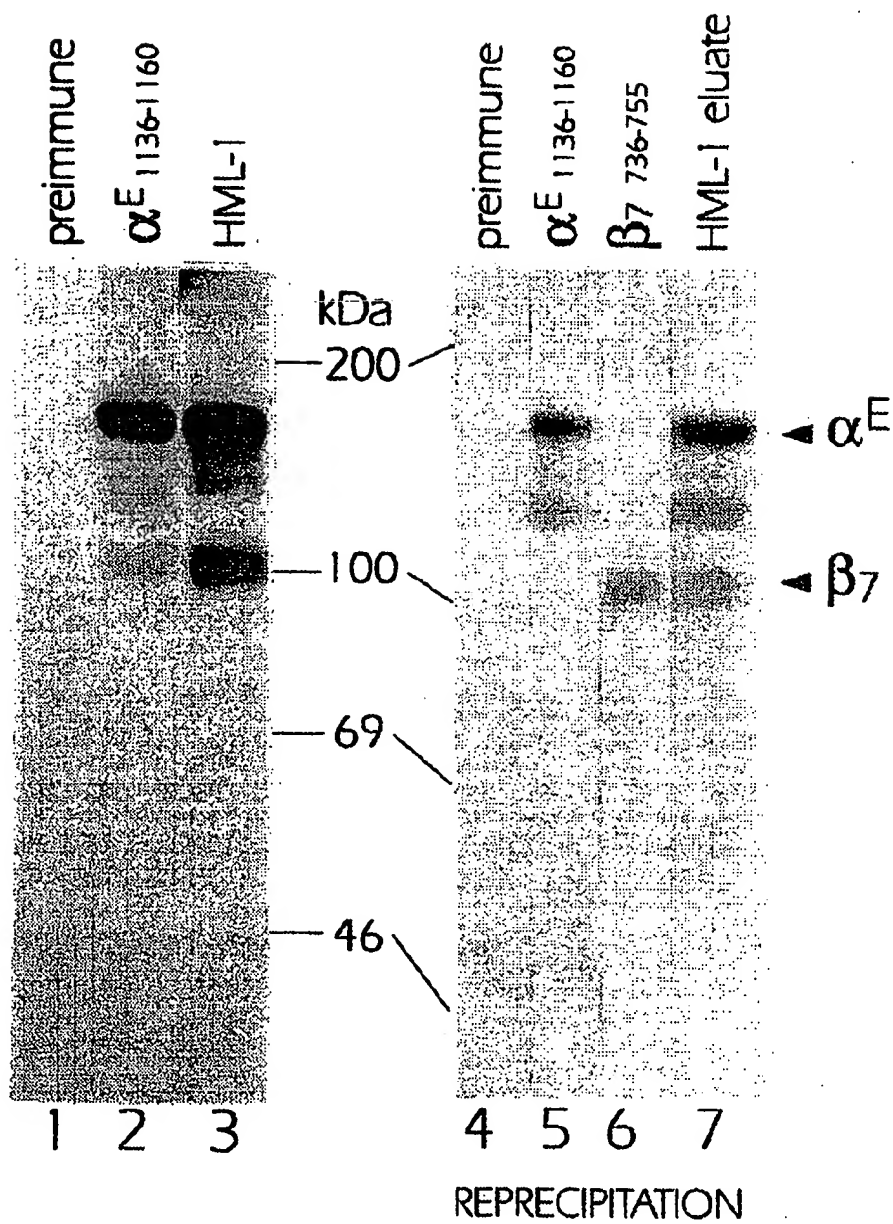


FIG. 4

FRESH LEUKOCYTE  
SUBPOPULATIONS

neutrophils  
PB-M  
PBL-T  
tonsillar B

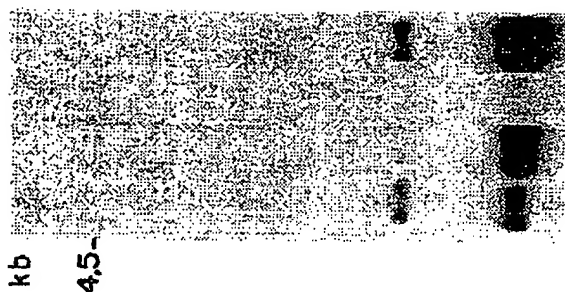


FIG. 5C

FRESH NORMAL TISSUES

heart  
brain  
placenta  
lung  
liver  
skeletal muscle  
kidney  
pancreas  
spleen  
thymus  
prostate  
testis  
ovary  
small intestine  
colon  
PB-leukocytes

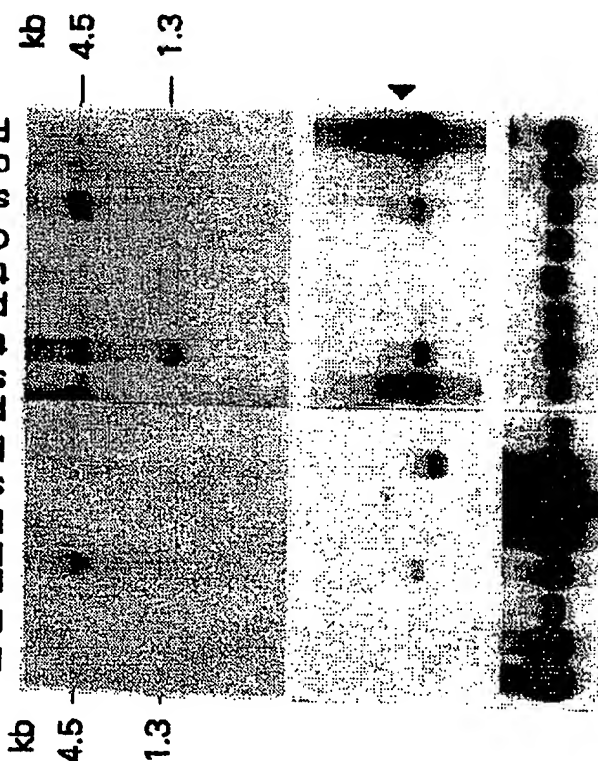


FIG. 5B

cultured iIEL  
hairy cell spleen

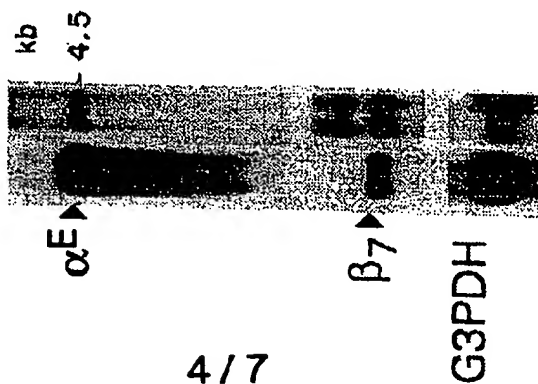


FIG. 5A

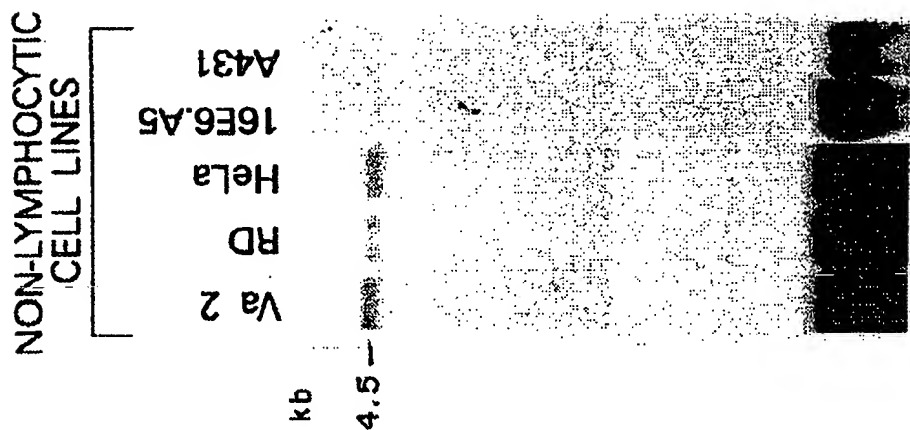


FIG. 5E

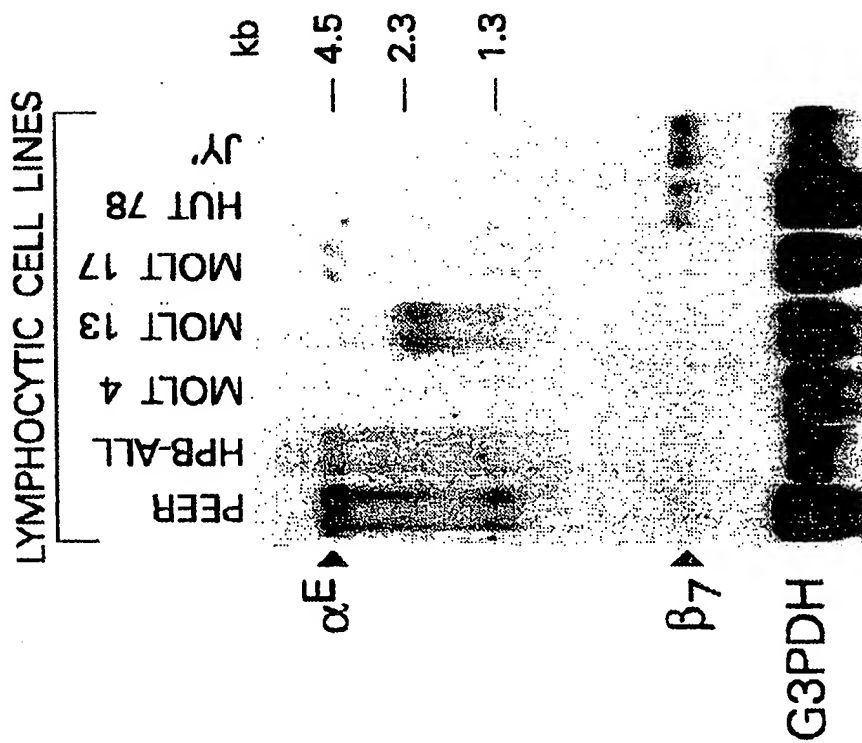


FIG. 5D

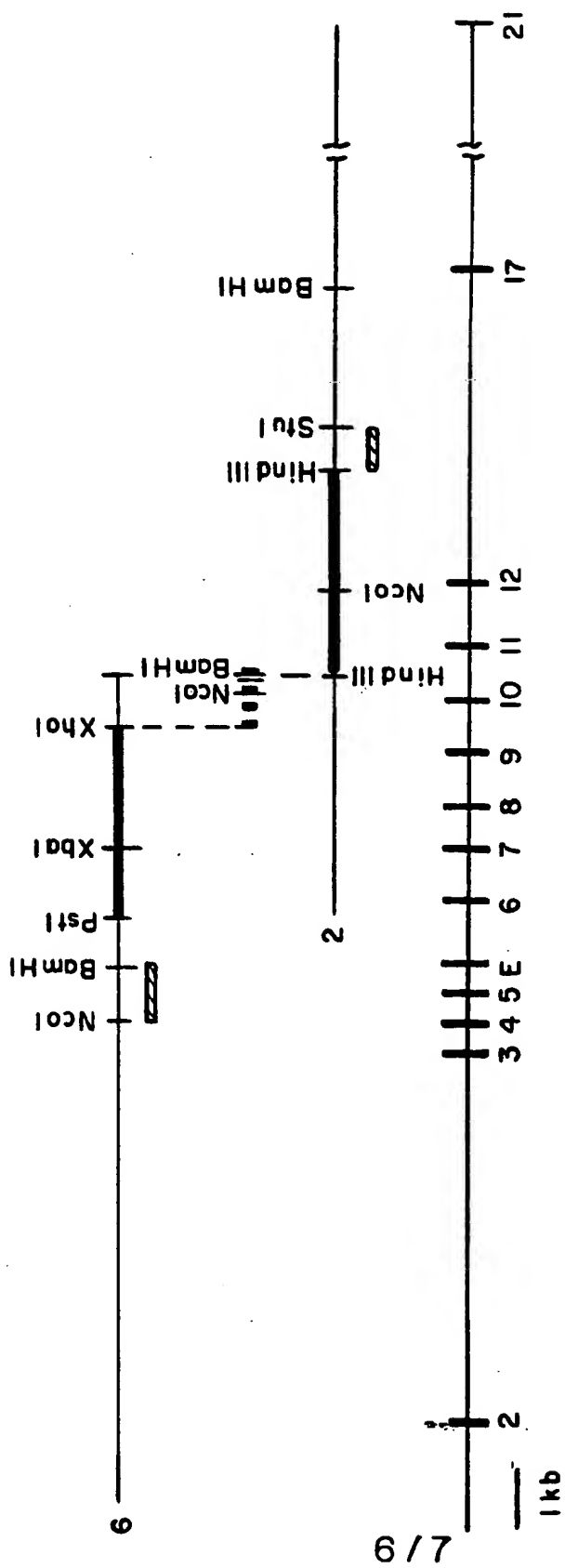


FIG.6



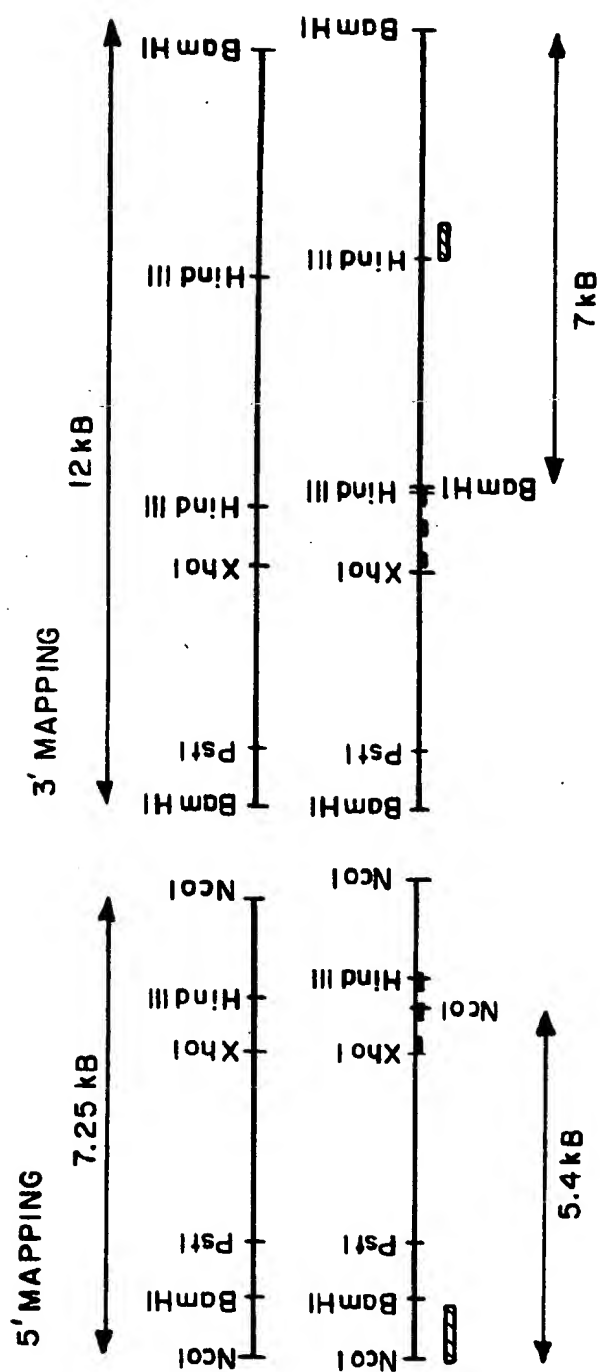


FIG. 7

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 95/02044

A. CLASSIFICATION OF SUBJECT MATTER  
 IPC 6 C12N15/12 C07K14/705 A61K38/17 G01N33/68 //C12N15/11,  
 A61K31/70

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C07K C12N A61K G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	PROC NATL ACAD SCI U S A 89 (5). 1992. 1924-1928, PARKER, C. ET AL. 'A FAMILY OF BETA-7 INTEGRINS ON HUMAN MUCOSAL LYMPHOCYTES.' cited in the application see the whole document ---	1,2,11
A	WO-A-91 07977 (SCRIPPS CLINIC RES) 13 June 1991 see page 21, line 14 - page 22, line 16 see page 34, line 26 - page 38 see examples ---	3-9
A	EUR J IMMUNOL 23 (7). 1993. 1630-1635, ROBERTS, K. ET AL. 'THE MUCOSAL T CELL INTEGRIN ALPHA -M-290-BETA-7 RECOGNIZES A LIGAND ON MUCOSAL EPITHELIAL CELL LINES.' see the whole document ---	11
-/--		

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

## \* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

14 June 1995

Date of mailing of the international search report

21-06-1995

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
 NL - 2280 HV Rijswijk  
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
 Fax (+31-70) 340-3016

Authorized officer

Andres, S

# INTERNATIONAL SEARCH REPORT

Inter mal Application No  
PCT/US 95/02044

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	<p>JOURNAL OF BIOLOGICAL CHEMISTRY 269 (8). 6016-6025, 25 February 1994 SHAW, S. ET AL. 'Molecular cloning of the human mucosal lymphocyte intergrin alpha-E subunit: Unusual structure and restricted RNA distribution.' see the whole document -----</p>	<p>12-14, 16-18</p>

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 95/02044

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9107977	13-06-91	US-A- 5196511	23-03-93
		AU-B- 645770	27-01-94
		AU-A- 6911391	26-06-91
		EP-A- 0502124	09-09-92
		JP-T- 5502228	22-04-93
		US-A- 5262520	16-11-93
-----			